Kennesaw State University
Department of Information Technology
IT6733 Database Administration
Spring 2018
Dr. Ying Xie

ETL Project:

2016 Presidential Election Data Using SSIS

By
Abraham Kim
Karis Kim

I. Data Source Link and Description

https://data.opendatasoft.com/explore/dataset/usa-2016-presidential-election-by-county@public/

OpenDataSoft

Contains 2016 presidential election results by county with additional demographic data as a csv and excel file. The csv file was downloaded for this project.

https://townhall.com/election/2016/president/

Townhall.com

Contains presidential election results from 2016 by each state then by each county in html format. The data in html source was scraped and transformed into a csv file using Python. Data was sorted by state in Townhall, but saved by county in our csv file.

http://www2.census.gov/geo/docs/reference/codes/files/national_county.txt (from https://www.census.gov/geo/reference/codes/cou.html)

U.S. Census Bureau

List of all states and its counties with state code, county code, FIPS code, and FIPS classification code as a txt file. FIPS(Federal Information Processing Standard) code is used to uniquely identify counties in the United States. The txt file was transformed into a csv file for this project.

https://catalog.data.gov/dataset/2016-general-election-results-by-precinct-complete-ecanvass-dataset/resource/76b36e87-aff3-47a8-86fd-9463d1551708

Data.gov

Contains 2016 general election results by each precinct. Precinct is the lowest level of division for election, also known as voting district. Presidential election results by precinct was extracted, but precinct data was determined to return a level of granularity that would be too fine for the purposes of this project, and therefore omitted from ETL. (Also, attempt to join by precinct above Data.gov source to OpenDataSoft failed, because precinct codes and names are not unique.) So, column “race” from Data.gov source was manually added to census destination table in SQL Server Data Tool step 37 below.

II. Justification of why a unified view on data from these sources is necessary

Presidential election data can provide valuable information, even after the votes have been tallied, the winner announced and sworn into office. An accurate and comprehensive/unified view of the election data can be used to affirm that the election process is accountable to public audit to ensure that fair, legal measures have indeed been upheld during the election. It is one of the cornerstones of democracy.

- OpenDataSoft source lists the USA 2016 Presidential Election by County with over a 100 columns of demographic data in addition to election results.

- Townhall source shows 2016 USA Presidential Election results by state and further by county and by party.

- The U.S. Census Bureau source lists all U.S. states and counties with respective FIPS code that can be used as a unique identifier.

- Data.gov source lists the results of the 2016 USA Presidential Election as well as other elections by precinct.

So, combining data from the Townhall source, the Census Bureau source and the Data.gov source can yield a unified view of the 2016 Presidential Election results by party, county and state, organized by a unique identifier, the FIPS code. Then with the OpenDataSoft source and its rich demographic information, the 2016 election results can be examined and analyzed with various demographic information through a query by joining the other combined data source with OpenDataSoft.

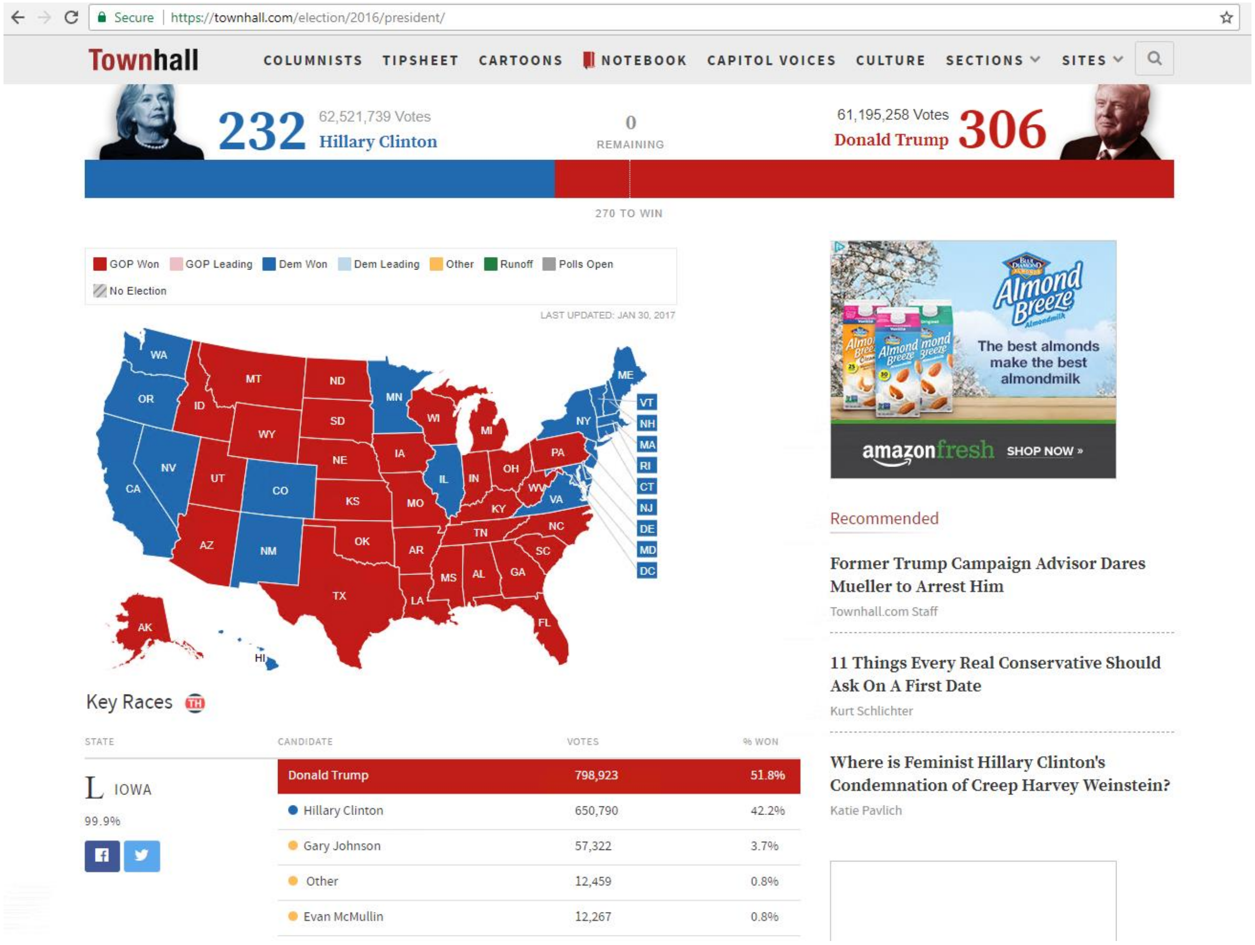
Having just the results of the presidential election (i.e.- total votes) would not help strategize the next election because it does not answer questions like how many males over females, what race, economic strata or geographic regions the voters came from. Also, not having a unique identifier like the FIPS code to ensure that there is no duplicate data, would lower the credibility of the studies from the data.

III. Instructions that will allow reviewer to recreate your project

A. Data Scraping

First, scraping the data from ‘https://townhall.com/election/2016/president’ with python codes.

First raw data is from html tags.



Key Races

STATE	CANDIDATE	VOTES	% WON
L IOWA 99.9%	Donald Trump	798,923	51.8%
	● Hillary Clinton	650,790	42.2%
	● Gary Johnson	57,322	3.7%
	● Other	12,459	0.8%
	● Evan McMullin	12,267	0.8%

Recommended

Former Trump Campaign Advisor Dares Mueller to Arrest Him
Townhall.com Staff

11 Things Every Real Conservative Should Ask On A First Date
Kurt Schlichter

Where is Feminist Hillary Clinton's Condemnation of Creep Harvey Weinstein?
Katie Pavlich

Below is Python code converting html to csv file:

import pandas as pd

import numpy as np

from bs4 import BeautifulSoup

import requests

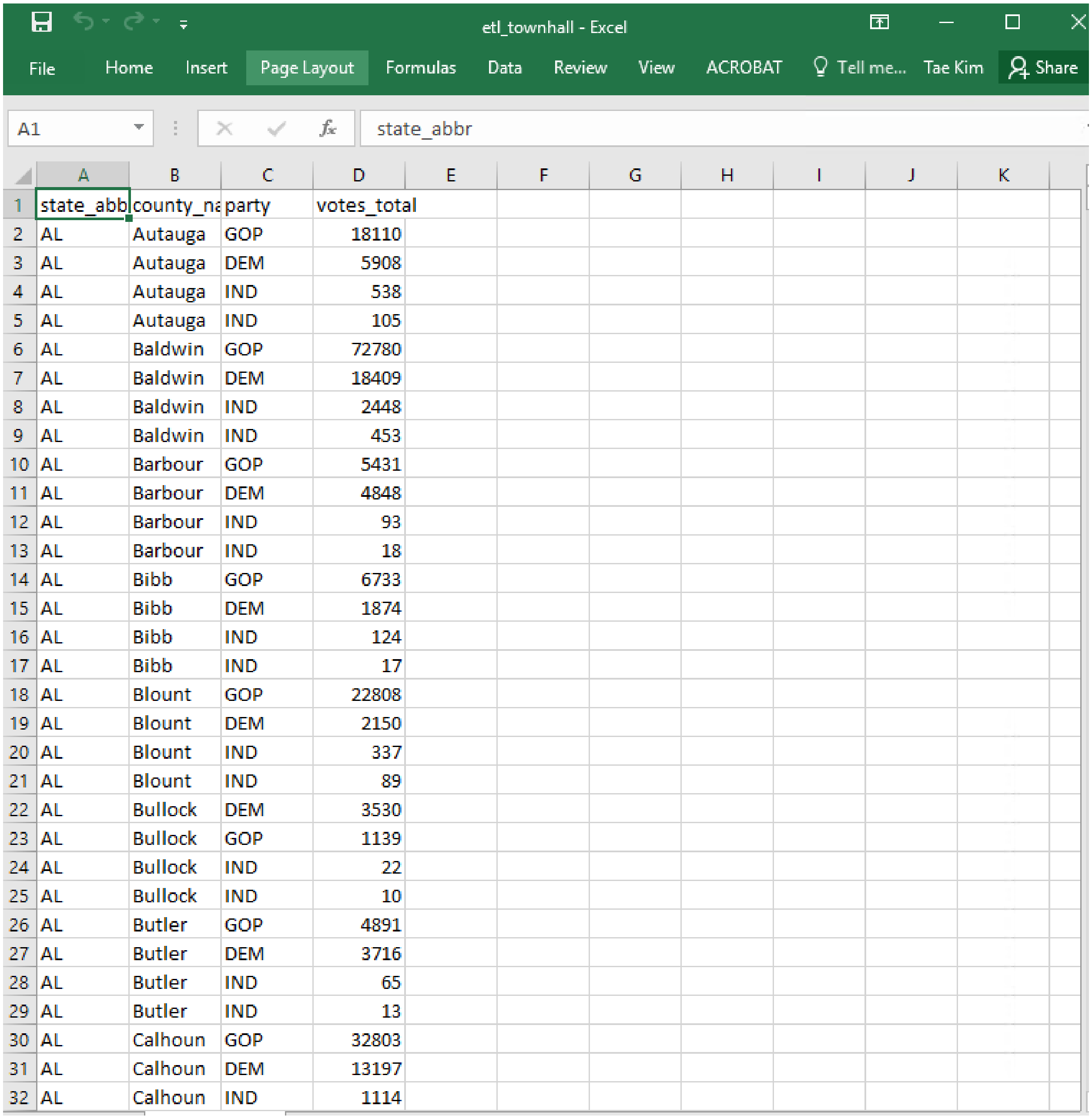
# Townhall data scraping

```
########################################################################################


# each page has a summary table that rolls up results at the state level
# get rid of it
def cond(x):
    if x:
        return x.startswith("table ec-table") and not "table ec-table ec-table-summary" in x
    else:
        return False

# list of state abbreviations
states = ['AL','AK','AZ','AR','CA','CO','CT','DC','DE','FL','GA','HI','ID','IL','IN',\
          'IA','KS','KY','LA','ME','MD','MA','MI','MN','MS','MO','MT','NE','NV','NH',\
          'NJ','NM','NY','NC','ND','OH','OK','OR','PA','RI','SC','SD','TN','TX','UT',\
          'VT','VA','WA','WV','WI','WY']

# headers for csv export
data = [['state_abbr', 'county_name', 'party', 'votes_total']]

# loop through each state's web page http://townhall.com/election/2016/president/%s/county,
where %s is the state abbr
for state in states:
    #r = req.urlopen('http://townhall.com/election/2016/president/' + state + '/county')
    page = requests.get('https://townhall.com/election/2016/president/' + state + '/county')
    soup = BeautifulSoup(page.text, 'html.parser')

    # loop through each <table> tag with .ec-table class
    tables = soup.findAll('table', attrs={'class':cond})
    for table in tables:
        if table.findParent("table") is None:
            table_body = table.find('tbody')

            rows = table_body.find_all('tr')
            for row in rows:
                cols = row.find_all('td')
                # first tbody tr has four td
                if len(cols) == 4:
                    # strip text from each td
                    divs = cols[0].find_all('div')
                    county = divs[0].text.strip()
                    party = cols[1]['class'][0]
                    total_votes = int(cols[2].text.strip().replace(',','').replace('-','0'))
                # all other tbody tr have three td
                else:
                    party = cols[1]['class'][0]
                    total_votes = int(cols[1].text.strip().replace(',','').replace('-','0'))

                #combine each row's results
                rowData = [state,county,party,total_votes]
                data.append(rowData)

townhall = pd.DataFrame(data) # throw results in dataframe
new_header = townhall.iloc[0] #grab the first row for the header
townhall = townhall[1:] #take the data less the header row
townhall.columns = new_header #set the header row as the df header
townhall['votes_total'] = townhall['votes_total'].astype('float64')
```

```
print(townhall.shape[0])
townhall.head()


print(townhall.loc[townhall['state_abbr'] =='DC', 'county_name'])
print(townhall.loc[townhall['county_name'] == 'Sainte Genevieve', 'county_name'])
print(townhall.loc[townhall['county_name'] == 'Oglala Lakota', 'county_name'])


# fix townhall county name for Washington DC, Sainte Genevieve, MO, Oglala, SD
townhall.loc[townhall['state_abbr'] =='DC', 'county_name'] = 'District of Columbia'
townhall.loc[townhall['county_name'] == 'Sainte Genevieve', 'county_name'] = 'Ste. Genevieve County'
townhall.loc[townhall['county_name'] == 'Oglala Lakota', 'county_name'] = 'Oglala'


print(townhall[(townhall['county_name'] == 'District of Columbia') |\
          (townhall['county_name'] == 'Ste. Genevieve County') |\
          (townhall['county_name'] == 'Oglala')])



# change 'Co.' to 'County' in county_name to match census county name
townhall['county_name'] = townhall['county_name'].apply(lambda x: x.replace('Co.','County').strip())
print(townhall[(townhall['state_abbr'] == 'NV') & (townhall['county_name'] == 'Carson City')])



# combine state and county names
#townhall['combined'] = townhall['state_abbr'] + townhall['county_name'].apply(lambda x: x.replace('
','').lower())
print(townhall[(townhall['state_abbr'] == 'NV') & (townhall['county_name'] == 'Carson City')])
print(townhall.shape)
townhall.head()
```

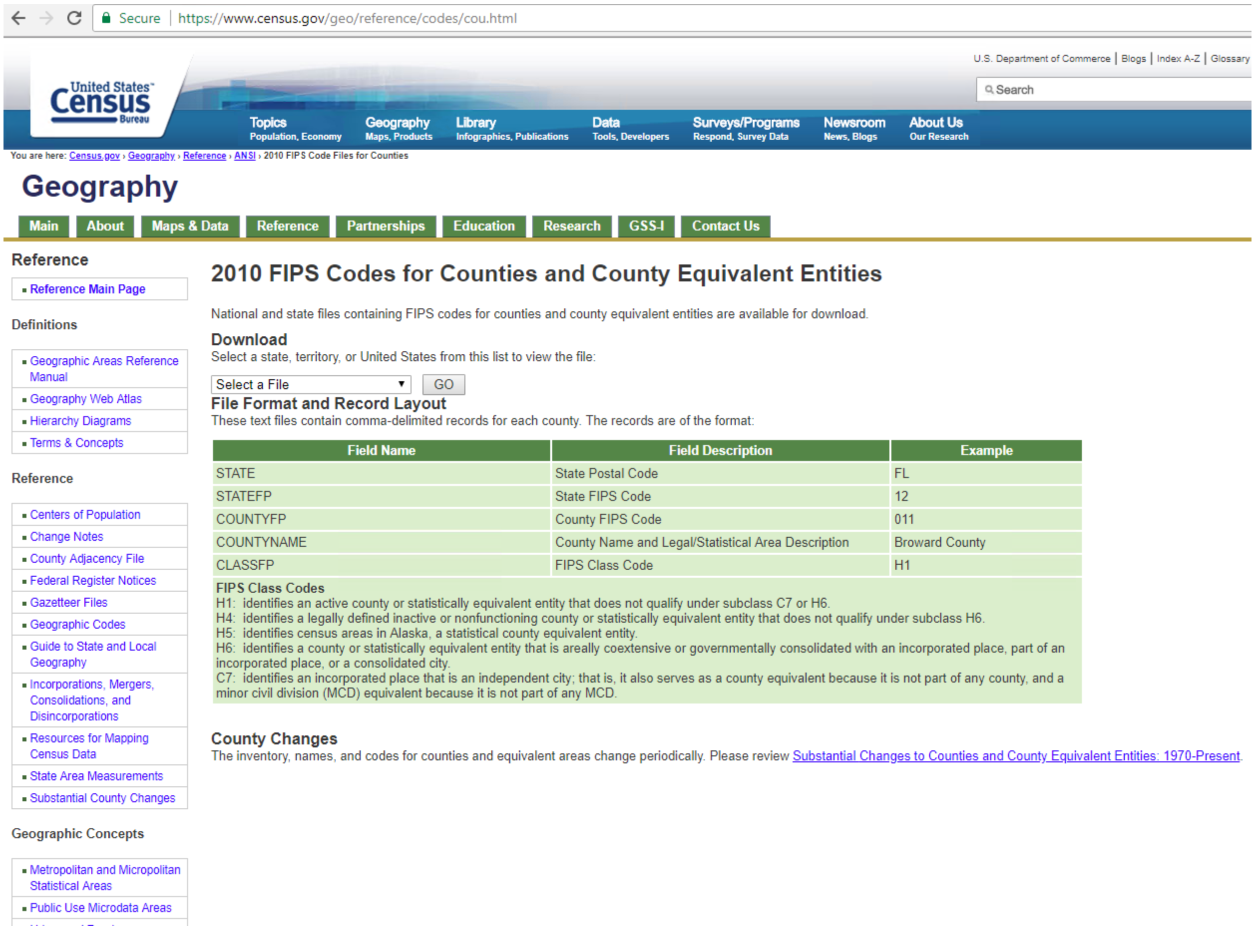
```
# send to csv files
townhall.to_csv('data/2016_presidential_election/etl_townhall.csv',sep=',',index=False)
```

Result:



	A	B	C	D
1	state_abb	county_na	party	votes_total
2	AL	Autauga	GOP	18110
3	AL	Autauga	DEM	5908
4	AL	Autauga	IND	538
5	AL	Autauga	IND	105
6	AL	Baldwin	GOP	72780
7	AL	Baldwin	DEM	18409
8	AL	Baldwin	IND	2448
9	AL	Baldwin	IND	453
10	AL	Barbour	GOP	5431
11	AL	Barbour	DEM	4848
12	AL	Barbour	IND	93
13	AL	Barbour	IND	18
14	AL	Bibb	GOP	6733
15	AL	Bibb	DEM	1874
16	AL	Bibb	IND	124
17	AL	Bibb	IND	17
18	AL	Blount	GOP	22808
19	AL	Blount	DEM	2150
20	AL	Blount	IND	337
21	AL	Blount	IND	89
22	AL	Bullock	DEM	3530
23	AL	Bullock	GOP	1139
24	AL	Bullock	IND	22
25	AL	Bullock	IND	10
26	AL	Butler	GOP	4891
27	AL	Butler	DEM	3716
28	AL	Butler	IND	65
29	AL	Butler	IND	13
30	AL	Calhoun	GOP	32803
31	AL	Calhoun	DEM	13197
32	AL	Calhoun	IND	1114

For Census data, scraping the data from ‘https://www.census.gov/geo/reference/codes/cou.html’ with python codes.


Secure
https://www.census.gov/geo/reference/codes/cou.html
U.S. Department of Commerce | Blogs | Index A-Z | Glossary
Search
United States Census Bureau
Topics Population, Economy
Geography Maps, Products
Library Infographics, Publications
Data Tools, Developers
Surveys/Programs Respond, Survey Data
Newsroom News, Blogs
About Us Our Research
You are here: Census.gov › Geography › Reference › ANSI › 2010 FIPS Code Files for Counties
Geography
Main
About
Maps & Data
Reference
Partnerships
Education
Research
GSS-I
Contact Us
Reference
Reference Main Page
Definitions
Geographic Areas Reference Manual
Geography Web Atlas
Hierarchy Diagrams
Terms & Concepts
Reference
Centers of Population
Change Notes
County Adjacency File
Federal Register Notices
Gazetteer Files
Geographic Codes
Guide to State and Local Geography
Incorporations, Mergers, Consolidations, and Disincorporations
Resources for Mapping Census Data
State Area Measurements
Substantial County Changes
Geographic Concepts
Metropolitan and Micropolitan Statistical Areas
Public Use Microdata Areas
Select a File
GO

2010 FIPS Codes for Counties and County Equivalent Entities

National and state files containing FIPS codes for counties and county equivalent entities are available for download.

Download

Select a state, territory, or United States from this list to view the file:

File Format and Record Layout

These text files contain comma-delimited records for each county. The records are of the format:

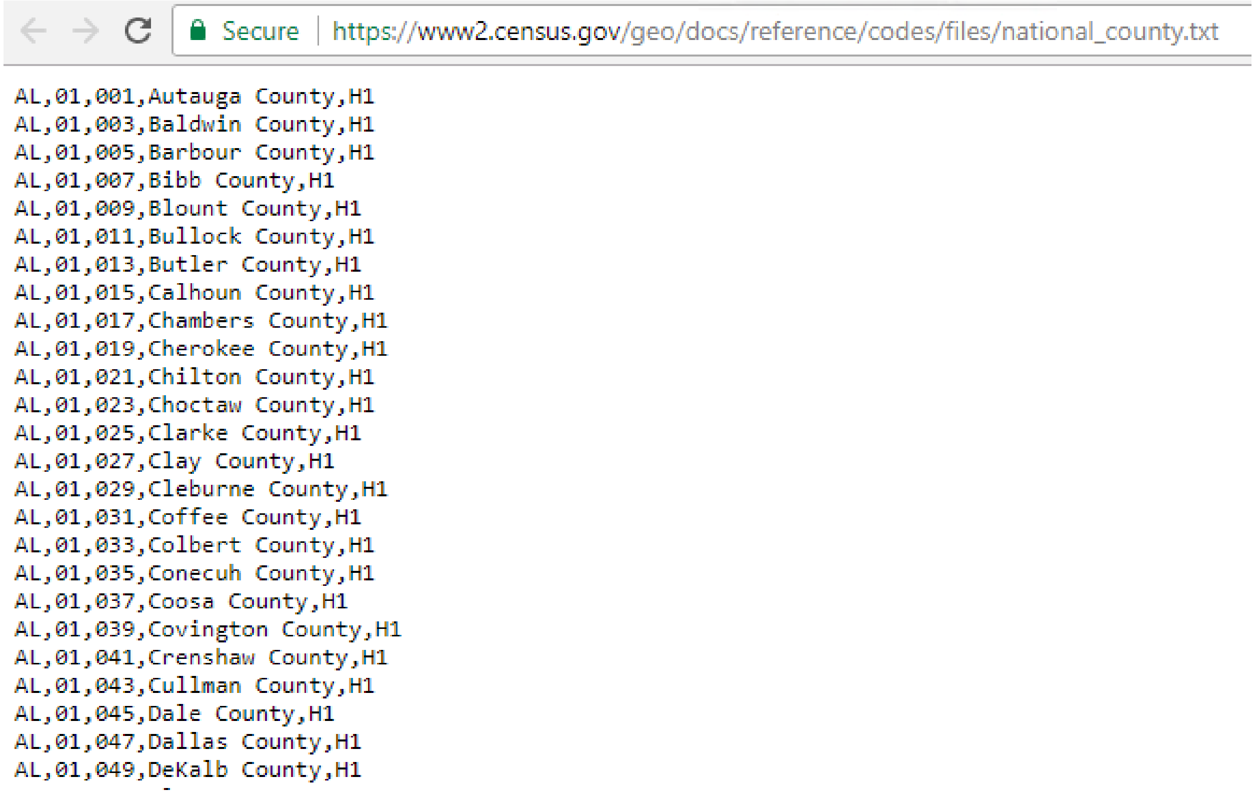
Field Name	Field Description	Example
STATE	State Postal Code	FL
STATEFP	State FIPS Code	12
COUNTYFP	County FIPS Code	011
COUNTYNAME	County Name and Legal/Statistical Area Description	Broward County
CLASSFP	FIPS Class Code	H1

FIPS Class Codes

H1: identifies an active county or statistically equivalent entity that does not qualify under subclass C7 or H6.
H4: identifies a legally defined inactive or nonfunctioning county or statistically equivalent entity that does not qualify under subclass H6.
H5: identifies census areas in Alaska, a statistical county equivalent entity.
H6: identifies a county or statistically equivalent entity that is areally coextensive or governmentally consolidated with an incorporated place, part of an incorporated place, or a consolidated city.
C7: identifies an incorporated place that is an independent city; that is, it also serves as a county equivalent because it is not part of any county, and a minor civil division (MCD) equivalent because it is not part of any MCD.

County Changes

The inventory, names, and codes for counties and equivalent areas change periodically. Please review Substantial Changes to Counties and County Equivalent Entities: 1970-Present.



Secure | https://www2.census.gov/geo/docs/reference/codes/files/national_county.txt

```
AL,01,001,Autauga County,H1
AL,01,003,Baldwin County,H1
AL,01,005,Barbour County,H1
AL,01,007,Bibb County,H1
AL,01,009,Blount County,H1
AL,01,011,Bullock County,H1
AL,01,013,Butler County,H1
AL,01,015,Calhoun County,H1
AL,01,017,Chambers County,H1
AL,01,019,Cherokee County,H1
AL,01,021,Chilton County,H1
AL,01,023,Choctaw County,H1
AL,01,025,Clarke County,H1
AL,01,027,Clay County,H1
AL,01,029,Cleburne County,H1
AL,01,031,Coffee County,H1
AL,01,033,Colbert County,H1
AL,01,035,Conecuh County,H1
AL,01,037,Coosa County,H1
AL,01,039,Covington County,H1
AL,01,041,Crenshaw County,H1
AL,01,043,Cullman County,H1
AL,01,045,Dale County,H1
AL,01,047,Dallas County,H1
AL,01,049,DeKalb County,H1
```

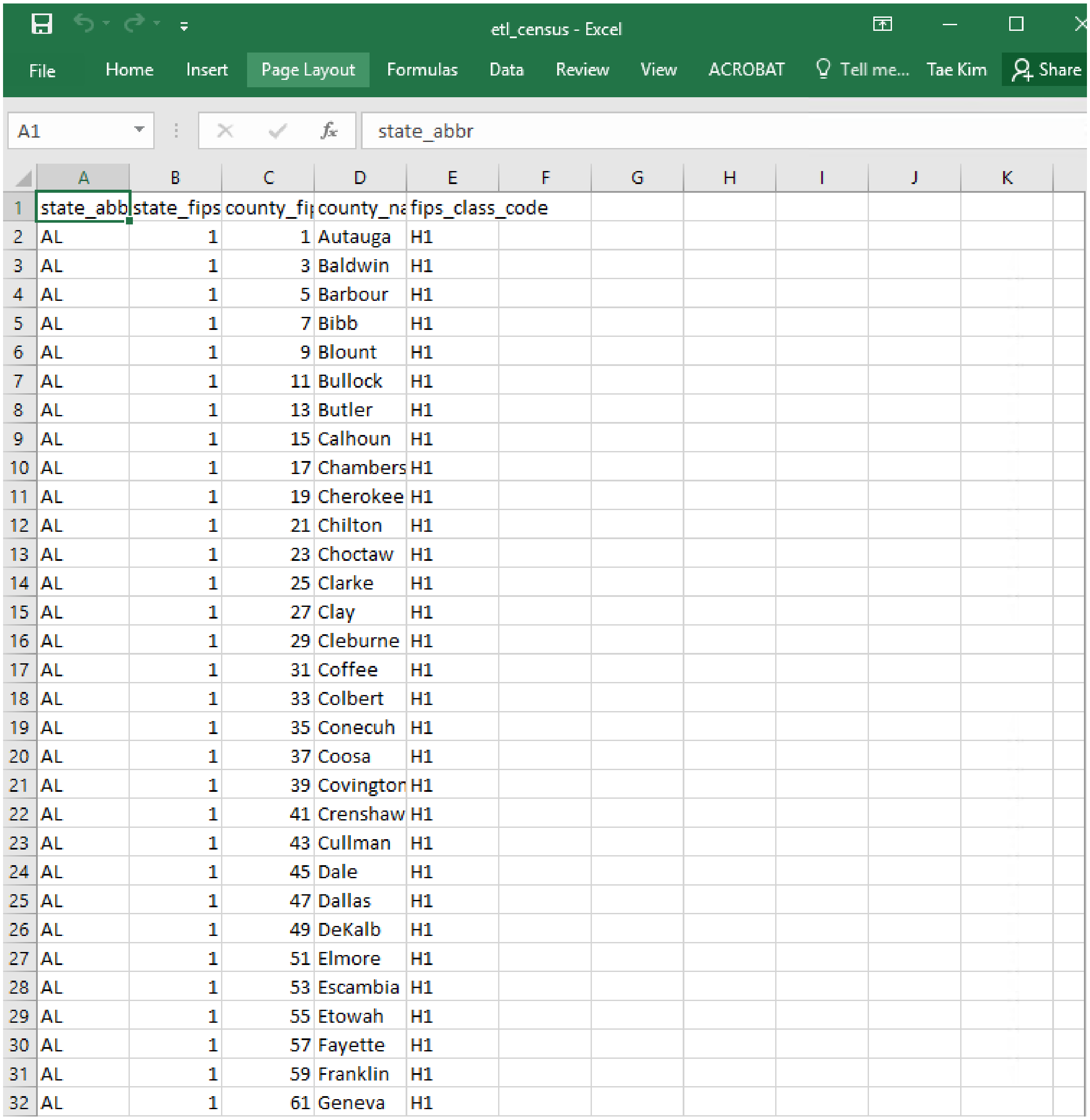
Python codes:

```
# Census data scraping

################################################################################


# county_fips data from https://www.census.gov/geo/reference/codes/cou.html

census = pd.read_csv('http://www2.census.gov/geo/docs/reference/codes/files/national_county.txt',sep=',',header=None, dtype=str)

census.columns = ['state_abbr', 'state_fips', 'county_fips', 'county_name', 'fips_class_code']

print(census.shape)

census.head()


# view by state

ak_counties = census[(census['state_abbr'] == 'AK')].shape[0]

print(ak_counties)

census[(census['state_abbr'] == 'AK')]
```

```
# change Shannon County, SD to Oglala County, SD
# http://rapidcityjournal.com/news/local/it-s-official-oglala-lakota-county-replaces-shannon-county-name/article_ac5c2369-3fea-5f94-9898-b007b7ddf22c.html
# townhall.loc[townhall['county_name'] == 'Sainte Genevieve', 'county_name'] = 'Ste. Genevieve County'
census.loc[(census['county_name'] == 'Shannon County') & (census['state_abbr'] == 'SD'), 'county_name'] = 'Oglala County'
census[(census['state_abbr'] == 'SD')]

# state of Alaska reports results at the precinct and state level; no county level data available
# report results as the states level;
# ugly fix to get townhall results and census counties to work together
# future plan: roll up precinct-level results to the county level
census.loc[(census['state_abbr'] == 'AK'),'county_name'] = 'Alaska'

# change county_name values in townhall data to match 'county_name' values for C7 fips class code cities
# get and transform C7 city names
cities = (census['state_abbr'][(census['fips_class_code'] == 'C7')] + census['county_name'][(census['fips_class_code'] == 'C7')]\
        .apply(lambda x: x.replace('city','').replace(' ','').lower()))
cities

# combine state and county names
townhall['combined'] = townhall['state_abbr'] + townhall['county_name'].apply(lambda x: x.replace(' ','').lower())

# loop through 'combined' column and compare to cities series to add 'city' to H1 fips class code to townhall data
for i, row in cities.iteritems():
    if row != 'NVcarsoncity':
```

```
    townhall.loc[townhall['combined'] == row, 'combined'] = row + 'city'

print(townhall[(townhall['combined'] == 'NVcarsoncity')])

# remove 'county' from 'combined' column of C7 fips class code counties in townhall
townhall['combined'] = townhall['combined'].str.replace('county','')
print(townhall[(townhall['county_name'] == 'Oglala')])

# return sum of votes by state and county
townhall['total_votes'] = townhall['votes_total'].groupby(townhall['combined']).transform('sum')
townhall_counties = townhall.drop('votes_total',axis=1)

# view dataset by selected state
print(townhall_counties[(townhall_counties['state_abbr'] == 'NV') & (townhall_counties['county_name'] == 'Carson City')])

census['county_name'] = census['county_name'].apply(lambda x: x.replace('County','').replace('Parish','').replace(' ',''))
print(census[(census['state_abbr'] == 'NV') & (census['county_name'] == 'Carson City')])
census.head()

census['fips'] = census['state_fips'] + census['county_fips']
print(census.head())

# send to csv files
#townhall.to_csv('data/2016_presidential_election/etl_townhall.csv',sep=',',index=False)
census.to_csv('data/2016_presidential_election/etl_census.csv',sep=',',index=False)
```

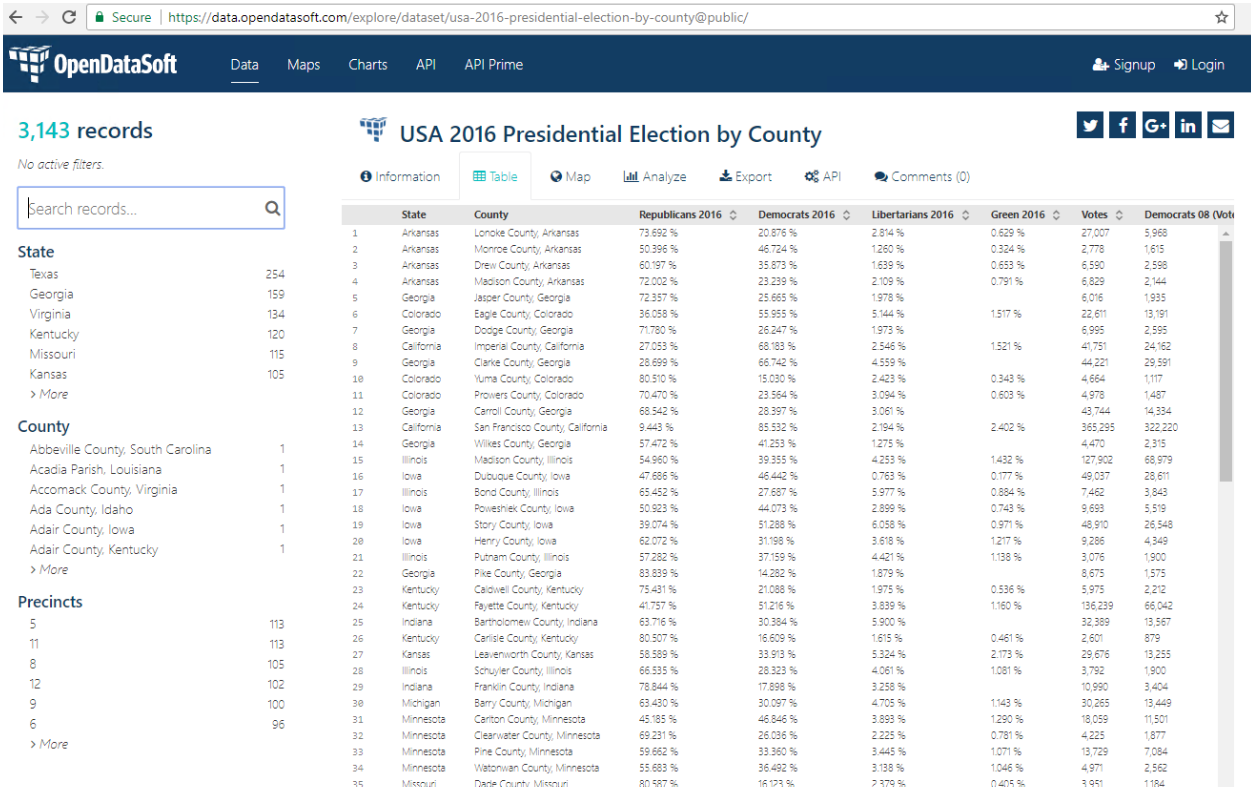
Result:



A1 | state_abbr

	A	B	C	D	E
1	state_abb	state_fips	county_fi	county_na	fips_class_code
2	AL	1	1	Autauga	H1
3	AL	1	3	Baldwin	H1
4	AL	1	5	Barbour	H1
5	AL	1	7	Bibb	H1
6	AL	1	9	Blount	H1
7	AL	1	11	Bullock	H1
8	AL	1	13	Butler	H1
9	AL	1	15	Calhoun	H1
10	AL	1	17	Chambers	H1
11	AL	1	19	Cherokee	H1
12	AL	1	21	Chilton	H1
13	AL	1	23	Choctaw	H1
14	AL	1	25	Clarke	H1
15	AL	1	27	Clay	H1
16	AL	1	29	Cleburne	H1
17	AL	1	31	Coffee	H1
18	AL	1	33	Colbert	H1
19	AL	1	35	Conecuh	H1
20	AL	1	37	Coosa	H1
21	AL	1	39	Covington	H1
22	AL	1	41	Crenshaw	H1
23	AL	1	43	Cullman	H1
24	AL	1	45	Dale	H1
25	AL	1	47	Dallas	H1
26	AL	1	49	DeKalb	H1
27	AL	1	51	Elmore	H1
28	AL	1	53	Escambia	H1
29	AL	1	55	Etowah	H1
30	AL	1	57	Fayette	H1
31	AL	1	59	Franklin	H1
32	AL	1	61	Geneva	H1

For Opendatasoft data, scraping the data from ‘https://data.opendatasoft.com/explore/dataset/usa-2016-presidential-election-by-county@public/’ with python codes.



3,143 records

No active filters.

USA 2016 Presidential Election by County

State

Texas	254
Georgia	159
Virginia	134
Kentucky	120
Missouri	115
Kansas	105

› *More*

County

Abbeville County, South Carolina	1
Acadia Parish, Louisiana	1
Accomack County, Virginia	1
Ada County, Idaho	1
Adair County, Iowa	1
Adair County, Kentucky	1

› *More*

Precincts

5	113
11	113
8	105
12	102
9	100
6	96

› *More*

	State	County	Republicans 2016	Democrats 2016	Libertarians 2016	Green 2016	Votes	Democrats 08 (Vote
1	Arkansas	Lonoke County, Arkansas	73.692 %	20.876 %	2.814 %	0.629 %	27,007	5,968
2	Arkansas	Monroe County, Arkansas	50.396 %	46.724 %	1.260 %	0.324 %	2,778	1,615
3	Arkansas	Drew County, Arkansas	60.197 %	35.873 %	1.639 %	0.653 %	6,590	2,598
4	Arkansas	Madison County, Arkansas	72.002 %	23.239 %	2.109 %	0.791 %	6,829	2,144
5	Georgia	Jasper County, Georgia	72.357 %	25.665 %	1.978 %		6,016	1,935
6	Colorado	Eagle County, Colorado	36.058 %	55.955 %	5.144 %	1.517 %	22,611	13,191
7	Georgia	Dodge County, Georgia	71.780 %	26.247 %	1.973 %		6,995	2,595
8	California	Imperial County, California	27.053 %	68.183 %	2.546 %	1.521 %	41,751	24,162
9	Georgia	Clarke County, Georgia	28.699 %	66.742 %	4.559 %		44,221	29,591
10	Colorado	Yuma County, Colorado	80.510 %	15.030 %	2.423 %	0.343 %	4,664	1,117
11	Colorado	Prowers County, Colorado	70.470 %	23.564 %	3.094 %	0.603 %	4,978	1,487
12	Georgia	Carroll County, Georgia	68.542 %	28.397 %	3.061 %		43,744	14,334
13	California	San Francisco County, California	9.443 %	85.532 %	2.194 %	2.402 %	365,295	322,220
14	Georgia	Wilkes County, Georgia	57.472 %	41.253 %	1.275 %		4,470	2,315
15	Illinois	Madison County, Illinois	54.960 %	39.355 %	4.253 %	1.432 %	127,902	68,979
16	Iowa	Dubuque County, Iowa	47.686 %	46.442 %	0.763 %	0.177 %	49,037	28,611
17	Illinois	Bond County, Illinois	65.452 %	27.687 %	5.977 %	0.884 %	7,462	3,843
18	Iowa	Poweshiek County, Iowa	50.923 %	44.073 %	2.899 %	0.743 %	9,693	5,519
19	Iowa	Story County, Iowa	39.074 %	51.288 %	6.058 %	0.971 %	48,910	26,548
20	Iowa	Henry County, Iowa	62.072 %	31.198 %	3.618 %	1.217 %	9,286	4,349
21	Illinois	Putnam County, Illinois	57.282 %	37.159 %	4.421 %	1.138 %	3,076	1,900
22	Georgia	Pike County, Georgia	83.839 %	14.282 %	1.879 %		8,675	1,575
23	Kentucky	Caldwell County, Kentucky	75.431 %	21.088 %	1.975 %	0.536 %	5,975	2,212
24	Kentucky	Fayette County, Kentucky	41.757 %	51.216 %	3.839 %	1.160 %	136,239	66,042
25	Indiana	Bartholomew County, Indiana	63.716 %	30.384 %	5.900 %		32,389	13,567
26	Kentucky	Carlisle County, Kentucky	80.507 %	16.609 %	1.615 %	0.461 %	2,601	879
27	Kansas	Leavenworth County, Kansas	58.589 %	33.913 %	5.324 %	2.173 %	29,676	13,255
28	Illinois	Schuyler County, Illinois	66.535 %	28.323 %	4.061 %	1.081 %	3,792	1,900
29	Indiana	Franklin County, Indiana	78.844 %	17.898 %	3.258 %		10,990	3,404
30	Michigan	Barry County, Michigan	63.430 %	30.097 %	4.705 %	1.143 %	30,265	13,449
31	Minnesota	Carlton County, Minnesota	45.185 %	46.846 %	3.893 %	1.290 %	18,059	11,501
32	Minnesota	Clearwater County, Minnesota	69.231 %	26.036 %	2.225 %	0.781 %	4,225	1,877
33	Minnesota	Pine County, Minnesota	59.662 %	33.360 %	3.445 %	1.071 %	13,729	7,084
34	Minnesota	Watonwan County, Minnesota	55.683 %	36.492 %	3.138 %	1.046 %	4,971	2,562
35	Missouri	Dade County, Missouri	80.587 %	16.123 %	2.379 %	0.405 %	3,951	1,184

Python Codes:

# Datagov data scraping

##

datagov = pd.read_csv('data/2016_presidential_election/2016_General_datagov_president.csv', sep=',', dtype=str)

print(datagov.shape)

datagov.tail(5)

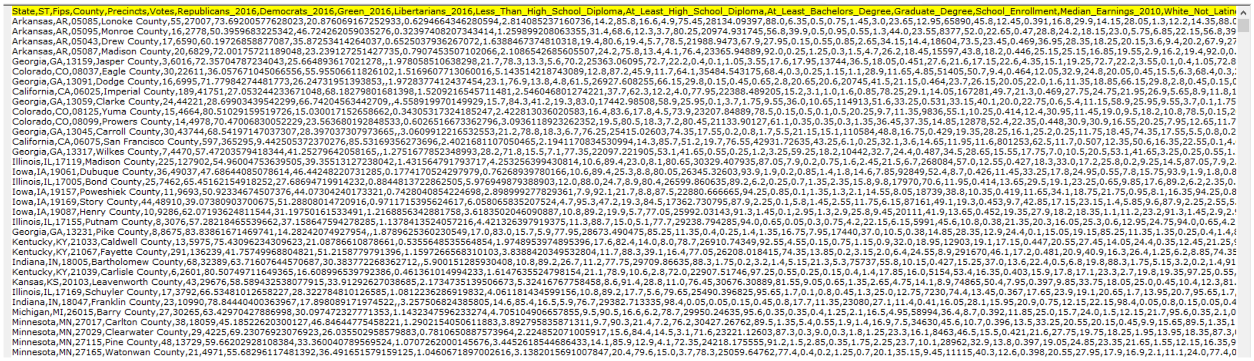
datagov_president = datagov.loc[(datagov['Race'] == 'US President & Vice President')]

print(datagov_president.shape)

datagov_president.head()

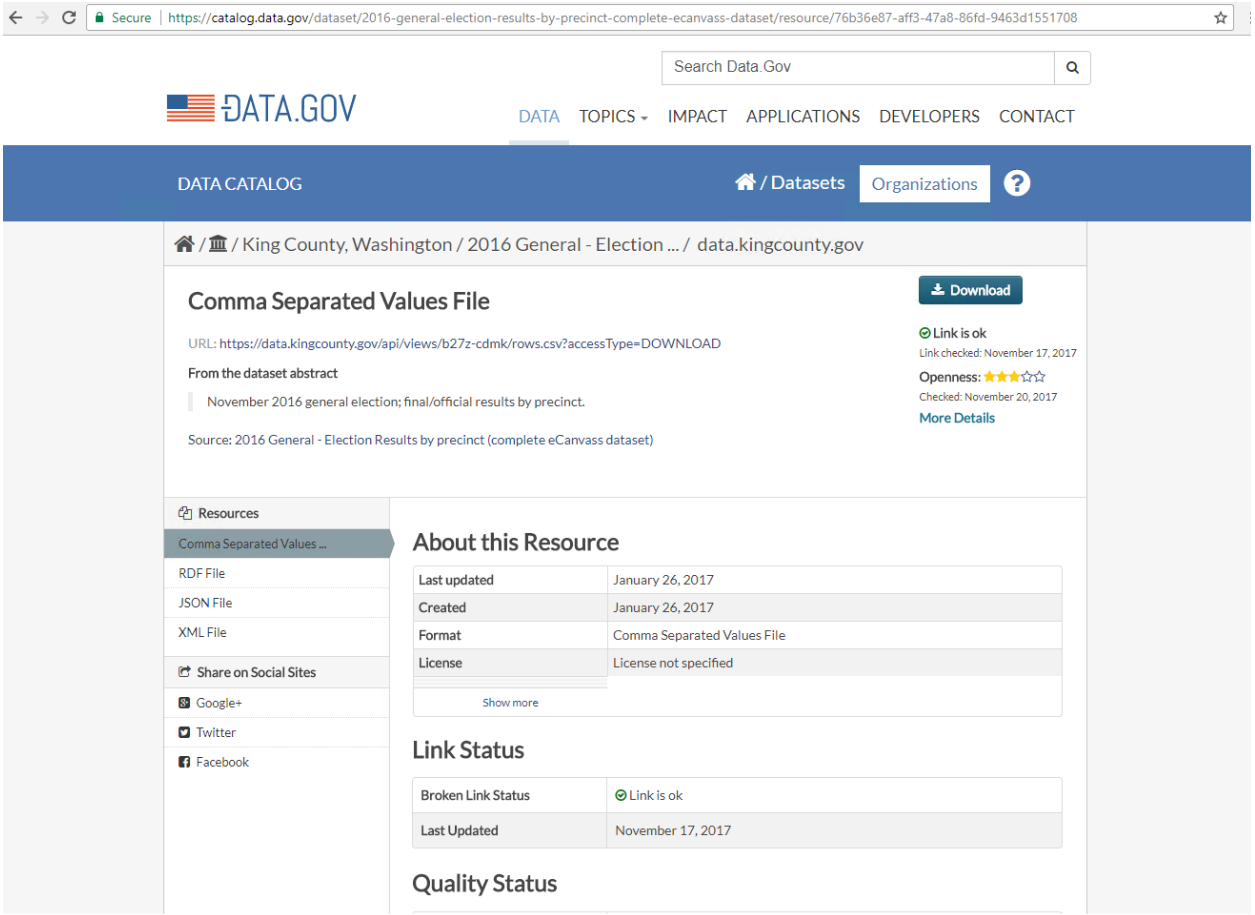
```
combined['Race'] = 'US President & Vice President'

combined.head()


# datagov_president to csv

datagov_president.to_csv('data/2016_presidential_election/etl_datagov.csv',sep=',', index=False)
```

Result:



```
State,ST,Fips,County,Precincts,Votes,Republicans_2016,Democrats_2016,Green_2016,Libertarians_2016,Less_Than_High_School_Diploma,At_Least_High_School_Diploma,At_Least_Bachelors_Degree,Graduate_Degree,School_Enrollment,Median_Earnings_2010,White_Not_Latin
Arkansas,AR,05085,Lonoke County,55,27007,73.69200577628023,20.876069167252933,0.6294664346280594,2.814085237160736,14.2,85.8,16.6,4.9,75.45,28134.09397,88.0,6.35,0.5,0.75,1.45,3.0,23.65,12.95,65890,45.8,12.45,0.391,16.8,29.9,14.15,28.05,1.3,12.2,14.35,88.0
Arkansas,AR,05095,Monroe County,16,2778,50.3959683225342,46.724262059035276,0.32397408207343414,1.259899208063355,31.4,68.6,12.3,3.7,80.25,20974.931745,56.8,39.9,0.5,0.95,0.55,1.3,44.0,23.55,8377,52.0,22.65,0.47,28.8,24.2,18.15,23.0,5.75,6.85,22.15,56.8,39
Arkansas,AR,05043,Drew County,17,6590,60.19726858877087,35.87253414264037,0.6525037936267072,1.6388467374810318,19.4,80.6,19.4,5.7,78.5,21988.9473,67.9,27.95,0.15,0.55,0.85,2.65,34.15,14.4,18604,73.5,23.45,0.469,36.95,28.35,18.25,20.15,3.6,9.4,20.2,67.9,27
Arkansas,AR,05087,Madison County,20,6829,72.00175721189048,23.239127251427735,0.7907453507102066,2.1086542685605507,24.2,75.8,13.4,4.1,76.4,23365.94889,92.0,0.25,1.25,0.3,1.5,4.7,26.2,18.45,15597,43.8,18.2,0.446,25.15,25.15,16.85,19.55,2.9,16.2,19.4,92.0,0.
Georgia,GA,13159,Jasper County,3,6016,72.35704787234043,25.664893617021278,,1.978058510638298,21.7,78.3,13.3,5.6,70.2,25363.06095,72.7,22.2,0.4,0.1,1.05,3.55,17.6,17.95,13744,36.5,18.05,0.451,27.6,21.6,17.15,22.6,4.35,15.1,19.25,72.7,22.2,3.55,0.1,0.4,1.05,72.8
Colorado,CO,08037,Eagle County,30,22611,36.057671045066556,55.95506611826102,1.5169607713060016,5.143514218743089,12.8,87.2,45.9,11.7,64.1,35484.543175,68.4,0.3,0.25,1.15,1.1,28.9,11.65,4.85,51405,50.7,9.4,0.464,12.05,32.9,24.8,20.05,0.45,15.5,6.3,68.4,0.3,
Georgia,GA,13091,Dodge County,16,6995,71.77984274481773,26.24731951393853,,1.9728377412437454,23.1,76.9,13.8,4.8,61.5,26927.608255,66.15,29.8,0.15,0.45,0.65,2.8,20.65,20.6,20745,41.5,21.15,0.464,23.7,26.15,20.05,22.0,1.6,11.35,18.85,66.15,29.8,2.8,0.45,0.15,0
California,CA,06025,Imperial County,189,41751,27.053244233671048,68.18279801681398,1.5209216545711481,2.546046801274221,37.7,62.3,12.2,4.0,77.95,22388.489205,15.2,3.1,1.0,1.6,0.85,78.25,29.1,14.05,167281,49.7,21.3,0.469,27.75,24.75,21.95,26.9,5.65,8.9,11.8,1
Georgia,GA,13059,Clarke County,24,44221,28.69903439542299,66.74204563442709,,4.558919970149929,15.7,84.3,41.2,19.3,83.0,17442.98508,58.9,25.95,0.1,3.7,1.75,9.55,36.0,10.65,114913,51.6,33.25,0.531,33.15,40.1,20.0,22.75,0.6,5.4,11.15,58.9,25.95,9.55,3.7,0.1,1.75
Colorado,CO,08125,Yuma County,15,4664,80.51029159519726,15.030017152658662,0.34305317324185247,2.422813036020583,16.4,83.6,17.8,4.5,73.9,23207.84889,78.5,0.15,0.5,0.1,0.5,20.25,9.7,11.35,9836,55.1,10.25,0.414,12.4,30.95,11.45,19.0,9.5,18.2,10.8,78.5,0.15,2
Colorado,CO,08099,Prowers County,14,4978,70.47006830052229,23.563680192848533,0.6026516673362796,3.0936118923262352,19.5,80.5,18.3,7.2,80.45,21133.90127,61.1,0.35,0.35,0.3,1.35,36.45,37.35,14.85,12878,52.4,22.35,0.448,30.9,30.9,16.55,20.25,7.95,12.65,11.7
Georgia,GA,13045,Carroll County,30,43744,68.54197147037307,28.397037307973665,,3.0609912216532553,21.2,78.8,18.3,6.7,76.25,25415.02603,74.35,17.55,0.2,0.8,1.7,5.5,21.15,15.1,110584,48.8,16.75,0.429,19.35,28.25,16.1,25.2,0.25,11.75,18.45,74.35,17.55,5.5,0.8,0.2
California,CA,06075,San Francisco County,597,365295,9.442505372370276,85.53169356273696,2.4021681107050465,2.1941170834530994,14.3,85.7,51.2,19.7,76.55,42931.72635,43.25,6.1,0.25,32.1,3.6,14.65,11.95,11.6,801253,62.5,11.7,0.507,12.35,50.6,16.35,22.55,0.1,4.
Georgia,GA,13317,Wilkes County,7,4470,57.47203579418344,41.25279642058165,,1.2751677852348993,28.2,71.8,15.5,7.1,77.35,22097.221905,53.1,41.65,0.55,0.25,1.2,3.25,59.25,18.2,10442,32.7,24.4,0.487,34.5,28.65,15.55,17.75,7.0,10.5,20.5,53.1,41.65,3.25,0.25,0.55,1.
Illinois,IL,17119,Madison County,225,127902,54.96004753639505,39.35513127238042,1.431564791793717,4.253256399430814,10.6,89.4,23.0,8.1,80.65,30329.407935,87.05,7.9,0.2,0.75,1.6,2.45,21.5,6.7,268084,57.0,12.55,0.427,18.3,33.0,17.2,25.8,0.2,9.25,14.5,87.05,7.9,2.
Iowa,IA,19061,Dubuque County,36,49037,47.68644085078614,46.44248220731285,0.1774170524297979,0.7626893978016 6,10.6,89.4,25.3,8.8,80.05,26345.32603,93.9,1.9,0.2,0.85,1.4,1.8,14.6,7.85,92849,52.4,8.7,0.426,11.45,33.25,17.8,24.95,0.55,7.8,15.75,93.9,1.9,1.8,0.8
Illinois,IL,17005,Bond County,25,7462,65.45162154918252,27.68694719914232,0.8844813722862505,5.976949879388903,12.0,88.0,24.7,8.9,80.4,26599.860635,89.2,6.2,0.25,0.7,1.35,2.35,15.8,9.8,17970,70.6,11.95,0.414,13.65,29.5,19.1,23.25,0.65,9.85,17.6,89.2,6.2,2.35,0.
Iowa,IA,19157,Poweshiek County,11,9693,50.92334674507376,44.07304240173321,0.7428040854224698,2.898999277829361,7.9,92.1,21.7,8.8,87.5,22880.666665,94.25,0.85,0.1,1.35,1.3,2.1,14.55,8.05,18739,38.8,10.35,0.419,11.65,34.1,18.75,21.75,0.95,8.1,16.35,94.25,0.8
Iowa,IA,19169,Story County,44,48910,39.07380903700675,51.28808014720916,0.9711715395624617,6.058065835207524,4.7,95.3,47.2,19.3,84.5,17362.730795,87.9,2.25,0.1,5.8,1.45,2.55,11.75,6.15,87161,49.1,19.3,0.453,9.7,42.85,17.15,23.15,1.4,5.85,9.6,87.9,2.25,2.55,5.
Iowa,IA,19087,Henry County,10,9286,62.07193624811544,31.19750161533491,1.2168856342881758,3.618350204609088 7,10.8,89.2,19.9,5.7,77.05,25992.03143,91.3,1.45,0.1,2.95,1.3,2.9,25.8,9.45,20111,41.9,13.65,0.452,19.35,27.9,18.2,18.35,1.1,11.2,23.2,91.3,1.45,2.9,2.
Illinois,IL,17155,Putnam County,8,3076,57.28218465539662,37.158647594278285,1.1378413524057216,4.421326397919375,11.3,88.7,15.0,5.1,77.7,29238.794285,94.0,0.65,0.05,0.3,0.75,4.2,22.15,6.15,5991,45.6,10.8,0.38,21.35,20.3,16.05,25.3,0.6,12.95,24.75,94.0,0.65,4.2
Georgia,GA,13231,Pike County,8,8675,83.83861671469741,14.28242074927954,,1.8789625360230549,17.0,83.0,15.7,5.9,77.95,28673.490475,85.25,11.35,0.4,0.25,1.4,1.35,16.75,7.95,17440,37.0,10.5,0.38,14.85,28.35,12.9,24.4,0.1,15.05,19.15,85.25,11.35,1.35,0.25,0.4,1.4,
Kentucky,KY,21033,Caldwell County,13,5975,75.43096234309623,21.08786610878661,0.5355648535564854,1.9748953974895396,17.6,82.4,14.0,8.0,78.7,26910.74349,92.55,4.55,0.15,0.75,1.15,0.9,32.0,18.95,12903,19.1,17.15,0.447,20.55,27.45,14.05,24.4,0.35,12.45,21.25,
Kentucky,KY,21067,Fayette County,291,136239,41.75749968804821,51.2158779791396,1.1597266568310103,3.8388420349532804,11.7,88.3,39.1,16.4,77.05,26208.018415,74.35,13.85,0.2,3.15,2.0,6.4,24.55,8.9,291670,46.1,17.2,0.481,20.9,40.9,16.3,26.4,1.25,6.2,8.85,74.35
Indiana,IN,18005,Bartholomew County,68,32389,63.71607644570687,30.383772268362712,,5.900151285930408,10.8,89.2,26.7,11.2,77.75,29709.86635,88.3,1.75,0.2,3.2,1.4,5.15,21.3,5.3,75737,55.8,10.15,0.427,15.25,37.0,13.6,22.4,0.5,6.8,19.8,88.3,1.75,5.15,3.2,0.2,1.4,91
Kentucky,KY,21039,Carlisle County,6,2601,80.50749711649365,16.608996539792386,0.461361014994233,1.6147635524798154,21.1,78.9,10.6,2.8,72.0,22907.51746,97.25,0.55,0.25,0.15,0.4,1.4,17.85,16.0,5154,53.4,16.35,0.403,15.9,17.8,17.1,23.3,2.7,19.8,19.35,97.25,0.55,
Kansas,KS,20103,Leavenworth County,43,29676,58.589432538077915,33.91292627038685,2.1734735139506673,5.32416767758458,8.6,91.4,28.8,11.0,76.45,30676.30889,81.55,9.05,0.65,1.35,2.65,4.75,14.1,8.9,74865,50.4,7.95,0.397,9.85,33.75,18.05,25.0,0.45,10.4,12.3,81.
Illinois,IL,17169,Schuyler County,17,3792,66.53481012658227,28.322784810126585,1.0812236286919832,4.061181434599156,10.8,89.2,17.7,5.6,79.65,25490.396825,95.65,1.7,0.1,0.8,0.45,1.3,25.0,12.75,7230,74.4,13.45,0.367,17.65,23.9,19.1,20.65,1.7,13.95,20.7,95.65,1.7,
Indiana,IN,18047,Franklin County,23,10990,78.84440400363967,17.898089171974522,,3.257506824385805,14.6,85.4,16.5,5.9,76.7,29382.713335,98.4,0.05,0.05,0.15,0.45,0.8,17.7,11.35,23080,27.1,11.4,0.41,16.05,28.1,15.95,20.9,0.75,12.15,22.15,98.4,0.05,0.8,0.15,0.05,0.4
Michigan,MI,26015,Barry County,27,30265,63.42970427886998,30.097472327771353,1.1432347596233274,4.705104906657855,9.5,90.5,16.6,6.2,78.7,29950.24635,95.6,0.35,0.35,0.4,1.25,2.1,16.5,4.95,58994,36.4,8.7,0.392,11.85,25.0,15.7,24.0,1.5,12.15,21.7,95.6,0.35,2.1,0
Minnesota,MN,27017,Carlton County,38,18059,45.18522620300127,46.84644775458221,1.2902154050611883,3.892795835871311,9.7,90.3,21.4,7.2,76.2,30427.26762,89.5,1.35,5.4,0.55,1.9,1.4,16.9,7.5,34630,45.6,10.7,0.396,13.5,33.25,20.55,20.15,0.45,9.9,15.65,89.5,1.35,1
Minnesota,MN,27029,Clearwater County,29,4225,69.23076923076923,26.035502958579883,0.7810650887573964,2.224852071005917,15.6,84.4,14.5,3.1,71.6,23221.12603,87.3,0.3,9.0,0.3,1.8,1.25,23.3,16.1,8463,46.5,15.5,0.421,21.6,27.75,19.75,18.25,1.95,13.95,18.35,87.3,
Minnesota,MN,27115,Pine County,48,13729,59.66202928108384,33.360040789569524,1.0707262000145676,3.4452618544686433,14.1,85.9,12.9,4.1,72.35,24218.175555,91.2,1.5,2.85,0.35,1.75,2.25,23.7,10.1,28962,32.9,13.8,0.397,19.05,24.85,23.35,21.65,1.55,12.15,16.35,9
Minnesota,MN,27165,Watonwan County,21,4971,55.68296117481392,36.491651579159125,1.0460671897002616,3.1382015691007847,20.4,79.6,15.0,3.7,78.3,25059.64762,77.4,0.4,0.2,1.25,0.7,20.1,35.15,9.45,11115,40.3,12.6,0.398,20.55,27.95,17.9,16.9,2.1,11.1,24.0,77.4,0
```

For Opendatasoft data, scraping the data from 'https://catalog.data.gov/dataset/2016-general-election-results-by-precinct-complete-ecanvass-dataset/resource/76b36e87-aff3-47a8-86fd-9463d1551708' with python codes.


Secure
https://catalog.data.gov/dataset/2016-general-election-results-by-precinct-complete-ecanvass-dataset/resource/76b36e87-aff3-47a8-86fd-9463d1551708
Search Data.Gov
DATA.GOV
DATA
TOPICS
IMPACT
APPLICATIONS
DEVELOPERS
CONTACT
DATA CATALOG
/ Datasets
Organizations
/ / King County, Washington / 2016 General - Election ... / data.kingcounty.gov
Download
Resources
Comma Separated Values ...
RDF File
JSON File
XML File
Share on Social Sites
Google+
Twitter
Facebook
Show more

Comma Separated Values File

URL: https://data.kingcounty.gov/api/views/b27z-cdmk/rows.csv?accessType=DOWNLOAD

From the dataset abstract

> November 2016 general election; final/official results by precinct.

Source: 2016 General - Election Results by precinct (complete eCanvass dataset)

✓ Link is ok
Link checked: November 17, 2017

Openness: ★★★☆☆
Checked: November 20, 2017

More Details

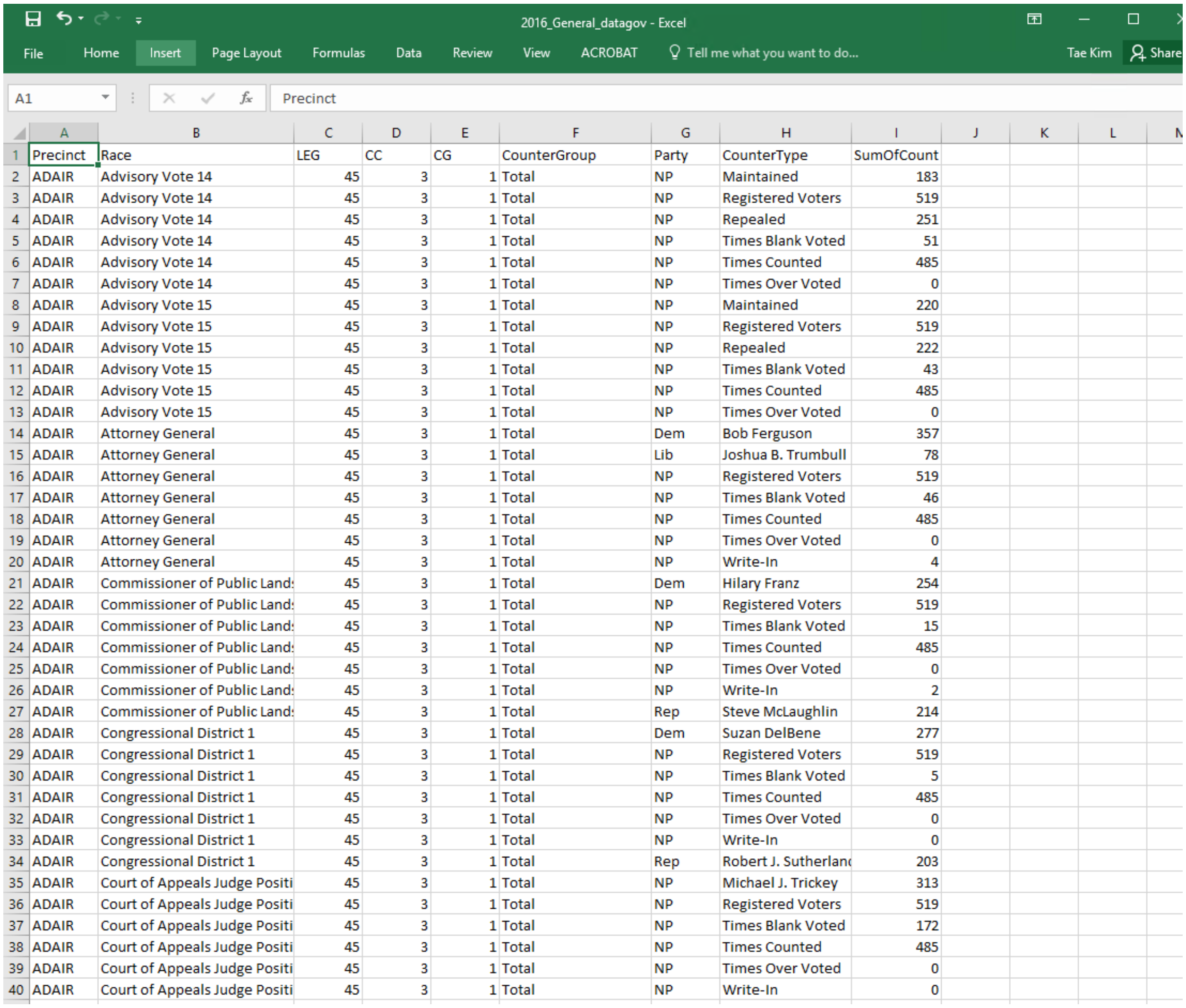
About this Resource

Last updated	January 26, 2017
Created	January 26, 2017
Format	Comma Separated Values File
License	License not specified

Link Status

Broken Link Status	✓ Link is ok
Last Updated	November 17, 2017

Quality Status



	A	B	C	D	E	F	G	H	I
1	Precinct	Race	LEG	CC	CG	CounterGroup	Party	CounterType	SumOfCount
2	ADAIR	Advisory Vote 14	45	3	1	Total	NP	Maintained	183
3	ADAIR	Advisory Vote 14	45	3	1	Total	NP	Registered Voters	519
4	ADAIR	Advisory Vote 14	45	3	1	Total	NP	Repealed	251
5	ADAIR	Advisory Vote 14	45	3	1	Total	NP	Times Blank Voted	51
6	ADAIR	Advisory Vote 14	45	3	1	Total	NP	Times Counted	485
7	ADAIR	Advisory Vote 14	45	3	1	Total	NP	Times Over Voted	0
8	ADAIR	Advisory Vote 15	45	3	1	Total	NP	Maintained	220
9	ADAIR	Advisory Vote 15	45	3	1	Total	NP	Registered Voters	519
10	ADAIR	Advisory Vote 15	45	3	1	Total	NP	Repealed	222
11	ADAIR	Advisory Vote 15	45	3	1	Total	NP	Times Blank Voted	43
12	ADAIR	Advisory Vote 15	45	3	1	Total	NP	Times Counted	485
13	ADAIR	Advisory Vote 15	45	3	1	Total	NP	Times Over Voted	0
14	ADAIR	Attorney General	45	3	1	Total	Dem	Bob Ferguson	357
15	ADAIR	Attorney General	45	3	1	Total	Lib	Joshua B. Trumbull	78
16	ADAIR	Attorney General	45	3	1	Total	NP	Registered Voters	519
17	ADAIR	Attorney General	45	3	1	Total	NP	Times Blank Voted	46
18	ADAIR	Attorney General	45	3	1	Total	NP	Times Counted	485
19	ADAIR	Attorney General	45	3	1	Total	NP	Times Over Voted	0
20	ADAIR	Attorney General	45	3	1	Total	NP	Write-In	4
21	ADAIR	Commissioner of Public Land	45	3	1	Total	Dem	Hilary Franz	254
22	ADAIR	Commissioner of Public Land	45	3	1	Total	NP	Registered Voters	519
23	ADAIR	Commissioner of Public Land	45	3	1	Total	NP	Times Blank Voted	15
24	ADAIR	Commissioner of Public Land	45	3	1	Total	NP	Times Counted	485
25	ADAIR	Commissioner of Public Land	45	3	1	Total	NP	Times Over Voted	0
26	ADAIR	Commissioner of Public Land	45	3	1	Total	NP	Write-In	2
27	ADAIR	Commissioner of Public Land	45	3	1	Total	Rep	Steve McLaughlin	214
28	ADAIR	Congressional District 1	45	3	1	Total	Dem	Suzan DelBene	277
29	ADAIR	Congressional District 1	45	3	1	Total	NP	Registered Voters	519
30	ADAIR	Congressional District 1	45	3	1	Total	NP	Times Blank Voted	5
31	ADAIR	Congressional District 1	45	3	1	Total	NP	Times Counted	485
32	ADAIR	Congressional District 1	45	3	1	Total	NP	Times Over Voted	0
33	ADAIR	Congressional District 1	45	3	1	Total	NP	Write-In	0
34	ADAIR	Congressional District 1	45	3	1	Total	Rep	Robert J. Sutherlan	203
35	ADAIR	Court of Appeals Judge Positi	45	3	1	Total	NP	Michael J. Trickey	313
36	ADAIR	Court of Appeals Judge Positi	45	3	1	Total	NP	Registered Voters	519
37	ADAIR	Court of Appeals Judge Positi	45	3	1	Total	NP	Times Blank Voted	172
38	ADAIR	Court of Appeals Judge Positi	45	3	1	Total	NP	Times Counted	485
39	ADAIR	Court of Appeals Judge Positi	45	3	1	Total	NP	Times Over Voted	0
40	ADAIR	Court of Appeals Judge Positi	45	3	1	Total	NP	Write-In	0

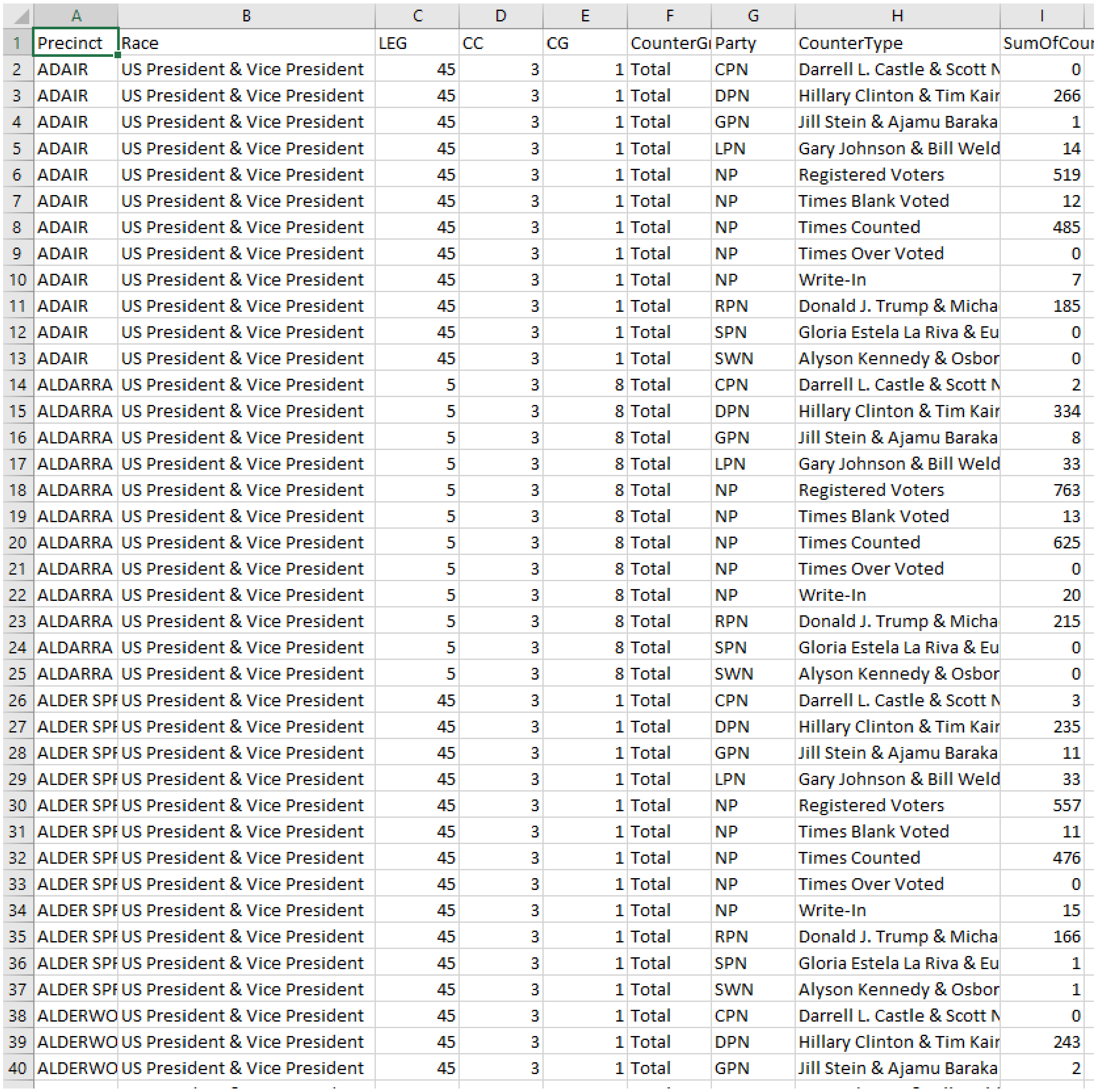
Extract only ‘US President & Vice President’ data from the dataset with Python.

Python Codes:

```
# Datagov data scraping

#####################################################################################

opendata = pd.read_csv('data/2016_presidential_election/2016-presidential-election-by-county-opendata.csv', sep=';', dtype=str)

print(opendata.shape)

opendata.head(5)
```

```
lst_col_opendata = opendata.columns
print(lst_col_opendata.values)

lst_mask_columns = ['Democrats 08 (Votes)', 'Democrats 12 (Votes)', 'Republicans 08 (Votes)',
'Republicans 12 (Votes)',
'Republicans 2012', 'Republicans 2008', 'Democrats 2012', 'Democrats 2008',
'total08', 'total12', 'other08', 'other12', 'other12_frac','other08_frac',
'rep12_frac2', 'rep08_frac2', 'dem12_frac2', 'dem08_frac2']

# df.loc[:, df.columns != 'b']
#opendata_extract = opendata[opendata.columns.difference(lst_mask_columns)]
opendata_extract = opendata.drop(opendata[(lst_mask_columns)], axis=1)
print(opendata_extract.shape)
opendata_extract.head()

# Cleansing columns name
#opendata_extract.columns = opendata_extract.columns.map(lambda x: int(x[1:]))
opendata_extract.columns = opendata_extract.columns.map(lambda x: \
   x.replace('(','').replace(' ','_').replace(')','').replace(',','_').replace("'s",'').replace('.','_'))
print(opendata_extract.columns.values)

# Reformat values for column a using an unnamed lambda function
opendata_extract['County'] = opendata_extract['County'].apply(lambda x: x.split(',')[0])
opendata_extract.head()

opendata_extract = opendata_extract.drop('statecode_prev', axis=1)
opendata_extract[opendata_extract['ST'] == 'GA']
len(combined[combined['state_abbr'] == 'GA'].fips.unique())
opendata_extract.to_csv('data/2016_presidential_election/etl_opendata.csv',sep=',', index=False)
```

Result:

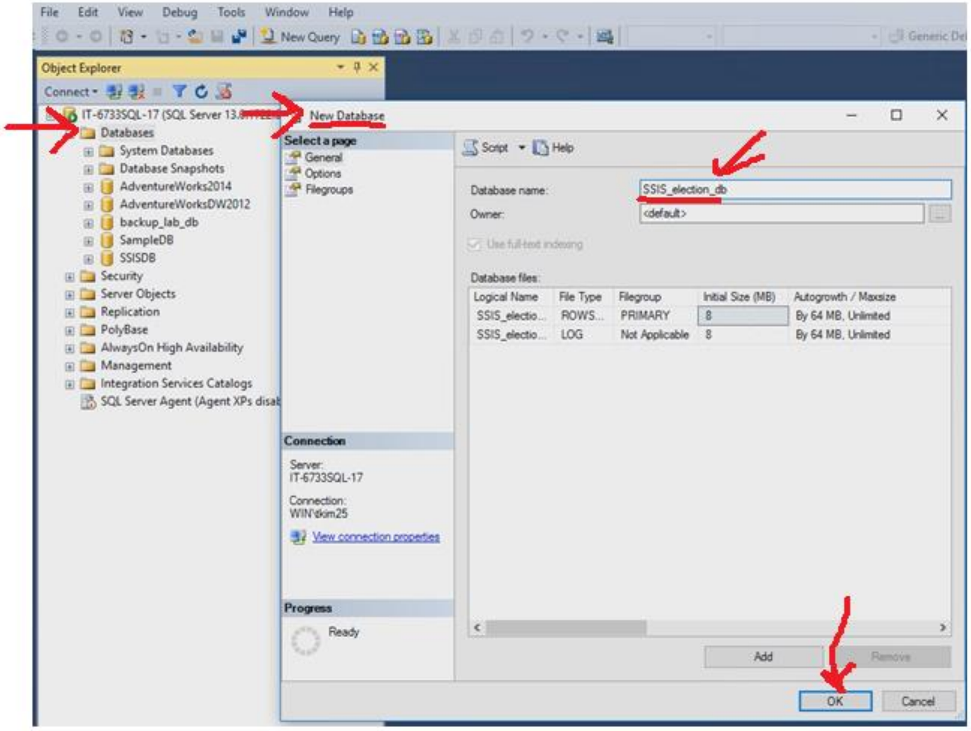


	A	B	C	D	E	F	G	H	I
1	Precinct	Race	LEG	CC	CG	CounterGr	Party	CounterType	SumOfCour
2	ADAIR	US President & Vice President	45	3	1	Total	CPN	Darrell L. Castle & Scott N	0
3	ADAIR	US President & Vice President	45	3	1	Total	DPN	Hillary Clinton & Tim Kair	266
4	ADAIR	US President & Vice President	45	3	1	Total	GPN	Jill Stein & Ajamu Baraka	1
5	ADAIR	US President & Vice President	45	3	1	Total	LPN	Gary Johnson & Bill Weld	14
6	ADAIR	US President & Vice President	45	3	1	Total	NP	Registered Voters	519
7	ADAIR	US President & Vice President	45	3	1	Total	NP	Times Blank Voted	12
8	ADAIR	US President & Vice President	45	3	1	Total	NP	Times Counted	485
9	ADAIR	US President & Vice President	45	3	1	Total	NP	Times Over Voted	0
10	ADAIR	US President & Vice President	45	3	1	Total	NP	Write-In	7
11	ADAIR	US President & Vice President	45	3	1	Total	RPN	Donald J. Trump & Micha	185
12	ADAIR	US President & Vice President	45	3	1	Total	SPN	Gloria Estela La Riva & Eu	0
13	ADAIR	US President & Vice President	45	3	1	Total	SWN	Alyson Kennedy & Osbor	0
14	ALDARRA	US President & Vice President	5	3	8	Total	CPN	Darrell L. Castle & Scott N	2
15	ALDARRA	US President & Vice President	5	3	8	Total	DPN	Hillary Clinton & Tim Kair	334
16	ALDARRA	US President & Vice President	5	3	8	Total	GPN	Jill Stein & Ajamu Baraka	8
17	ALDARRA	US President & Vice President	5	3	8	Total	LPN	Gary Johnson & Bill Weld	33
18	ALDARRA	US President & Vice President	5	3	8	Total	NP	Registered Voters	763
19	ALDARRA	US President & Vice President	5	3	8	Total	NP	Times Blank Voted	13
20	ALDARRA	US President & Vice President	5	3	8	Total	NP	Times Counted	625
21	ALDARRA	US President & Vice President	5	3	8	Total	NP	Times Over Voted	0
22	ALDARRA	US President & Vice President	5	3	8	Total	NP	Write-In	20
23	ALDARRA	US President & Vice President	5	3	8	Total	RPN	Donald J. Trump & Micha	215
24	ALDARRA	US President & Vice President	5	3	8	Total	SPN	Gloria Estela La Riva & Eu	0
25	ALDARRA	US President & Vice President	5	3	8	Total	SWN	Alyson Kennedy & Osbor	0
26	ALDER SPF	US President & Vice President	45	3	1	Total	CPN	Darrell L. Castle & Scott N	3
27	ALDER SPF	US President & Vice President	45	3	1	Total	DPN	Hillary Clinton & Tim Kair	235
28	ALDER SPF	US President & Vice President	45	3	1	Total	GPN	Jill Stein & Ajamu Baraka	11
29	ALDER SPF	US President & Vice President	45	3	1	Total	LPN	Gary Johnson & Bill Weld	33
30	ALDER SPF	US President & Vice President	45	3	1	Total	NP	Registered Voters	557
31	ALDER SPF	US President & Vice President	45	3	1	Total	NP	Times Blank Voted	11
32	ALDER SPF	US President & Vice President	45	3	1	Total	NP	Times Counted	476
33	ALDER SPF	US President & Vice President	45	3	1	Total	NP	Times Over Voted	0
34	ALDER SPF	US President & Vice President	45	3	1	Total	NP	Write-In	15
35	ALDER SPF	US President & Vice President	45	3	1	Total	RPN	Donald J. Trump & Micha	166
36	ALDER SPF	US President & Vice President	45	3	1	Total	SPN	Gloria Estela La Riva & Eu	1
37	ALDER SPF	US President & Vice President	45	3	1	Total	SWN	Alyson Kennedy & Osbor	1
38	ALDERWO	US President & Vice President	45	3	1	Total	CPN	Darrell L. Castle & Scott N	0
39	ALDERWO	US President & Vice President	45	3	1	Total	DPN	Hillary Clinton & Tim Kair	243
40	ALDERWO	US President & Vice President	45	3	1	Total	GPN	Jill Stein & Ajamu Baraka	2

Attempt to join by precinct above Data.gov source to OpenDataSoft failed, because precinct codes and names are not unique. So, column “race” was added to census destination table in SQL Server Data Tool.

B. Build & Deploy

1. In SQL Server Management Studio, create a DB container for the ETL project.



Database name: SSIS_election_db

Owner: <default>

☑ Use full-text indexing

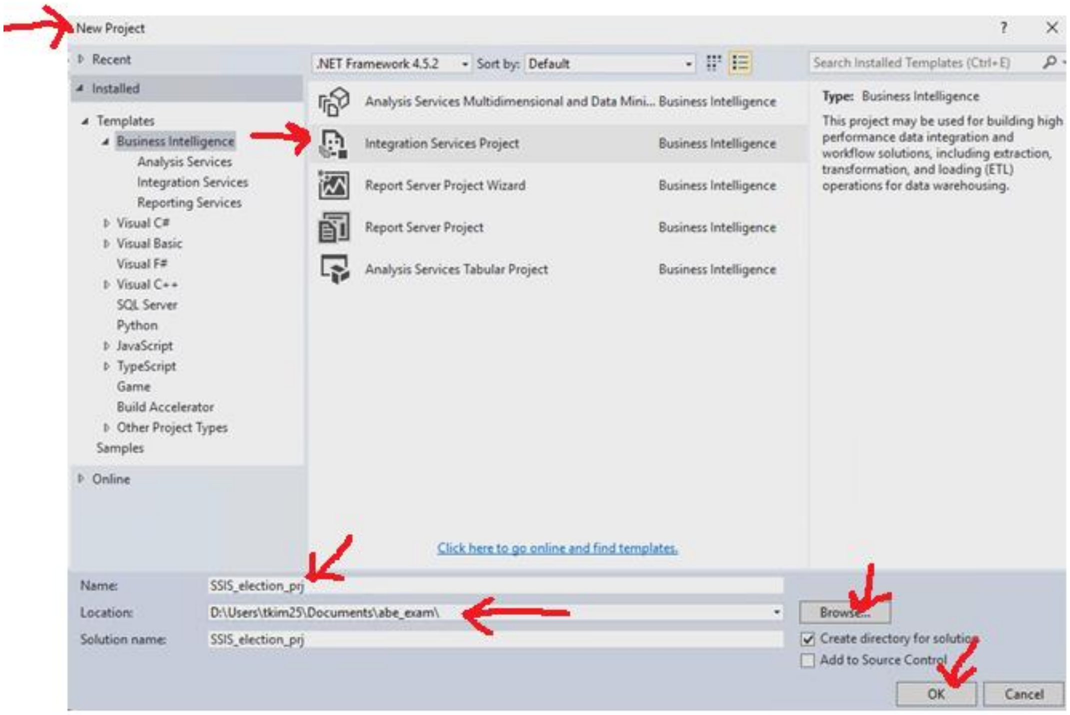
Server: IT-6733SQL-17

Connection: WIN\tkim25

Database files:

Logical Name	File Type	Filegroup	Initial Size (MB)	Autogrowth / Maxsize
SSIS_electio...	ROWS...	PRIMARY	8	By 64 MB, Unlimited
SSIS_electio...	LOG	Not Applicable	8	By 64 MB, Unlimited

2. In SQL Server Data Tool 2015, create a new project.


New Project
Recent
Installed
Templates
Business Intelligence
Analysis Services
Integration Services
Reporting Services
Visual C#
Visual Basic
Visual F#
Visual C++
SQL Server
Python
JavaScript
TypeScript
Game
Build Accelerator
Other Project Types
Samples
Online
.NET Framework 4.5.2
Sort by: Default
Search Installed Templates (Ctrl+E)
Analysis Services Multidimensional and Data Mini... Business Intelligence
Integration Services Project Business Intelligence
Report Server Project Wizard Business Intelligence
Report Server Project Business Intelligence
Analysis Services Tabular Project Business Intelligence
Click here to go online and find templates.
Browse...
OK
Cancel

Type: Business Intelligence

This project may be used for building high performance data integration and workflow solutions, including extraction, transformation, and loading (ETL) operations for data warehousing.

Name: SSIS_election_prj

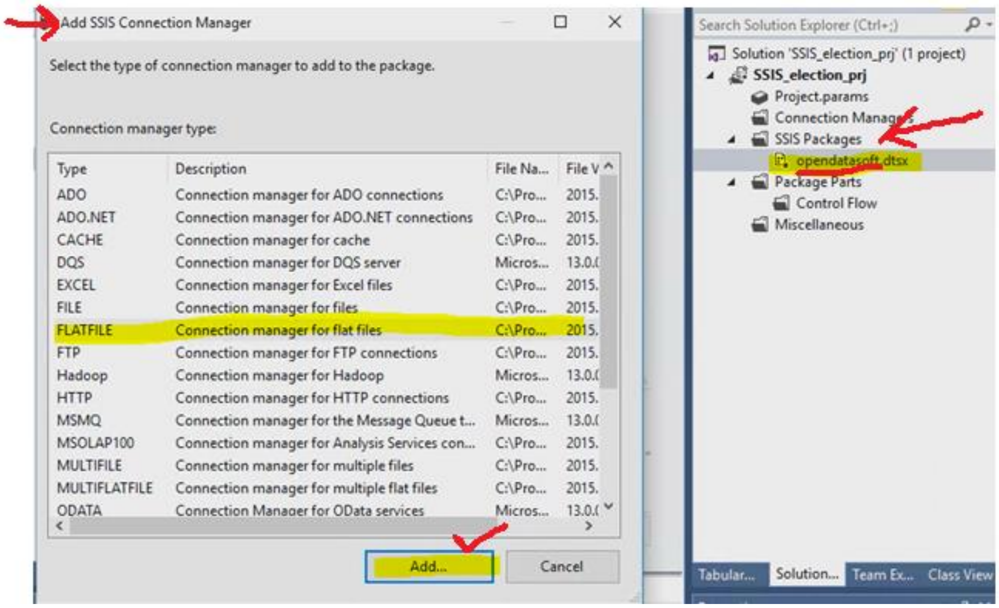
Location: D:\Users\tkim25\Documents\abe_exam\

Solution name: SSIS_election_prj

☑ Create directory for solution

☐ Add to Source Control

3. Create .dtsx files for our data sources:
   Under *SSIS Packages* on the right, rename project. In Control Flow tab, mouse right click on design surface in the middle, select New Connection, then in Add SSIS Connection Manager, select FLATFILE connection manager type and Add.



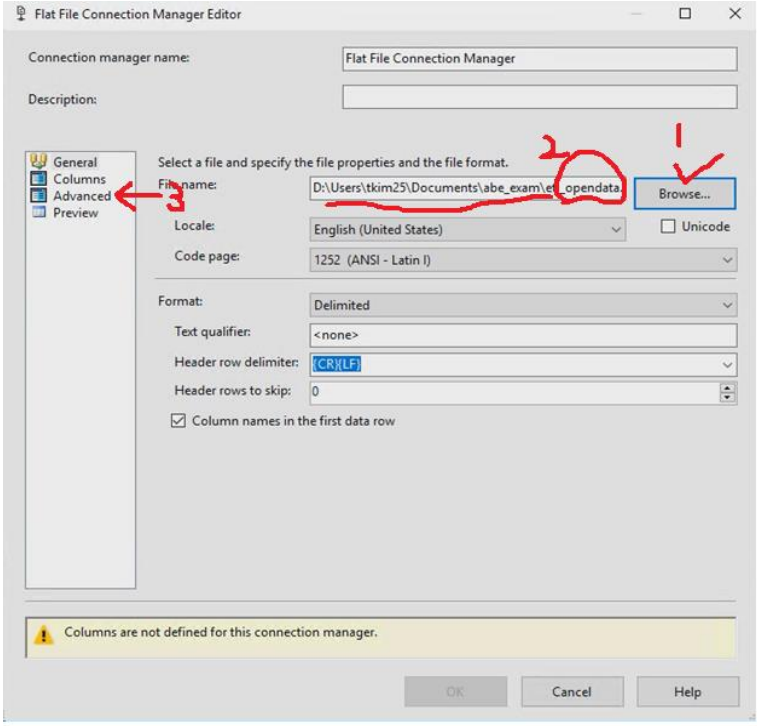
Add SSIS Connection Manager

Select the type of connection manager to add to the package.

Connection manager type:

Type	Description	File Na...	File V
ADO	Connection manager for ADO connections	C:\Pro...	2015.
ADO.NET	Connection manager for ADO.NET connections	C:\Pro...	2015.
CACHE	Connection manager for cache	C:\Pro...	2015.
DQS	Connection manager for DQS server	Micros...	13.0.(
EXCEL	Connection manager for Excel files	C:\Pro...	2015.
FILE	Connection manager for files	C:\Pro...	2015.
FLATFILE	Connection manager for flat files	C:\Pro...	2015.
FTP	Connection manager for FTP connections	C:\Pro...	2015.
Hadoop	Connection manager for Hadoop	Micros...	13.0.(
HTTP	Connection manager for HTTP connections	C:\Pro...	2015.
MSMQ	Connection manager for the Message Queue t...	Micros...	13.0.(
MSOLAP100	Connection manager for Analysis Services con...	C:\Pro...	2015.
MULTIFILE	Connection manager for multiple files	C:\Pro...	2015.
MULTIFLATFILE	Connection manager for multiple flat files	C:\Pro...	2015.
ODATA	Connection Manager for OData services	Micros...	13.0.(

4. In the next window, click Browse, then in the next Open file window select/open the folder with my data source files, select CSV files in lower right to see those files. For this, select OpenData.csv, then click Open.



Flat File Connection Manager Editor

Connection manager name: Flat File Connection Manager

Description:

Select a file and specify the file properties and the file format.

File name: D:\Users\tkim25\Documents\abe_exam\e_opendata.

Locale: English (United States)

☐ Unicode

Code page: 1252 (ANSI - Latin I)

Format: Delimited

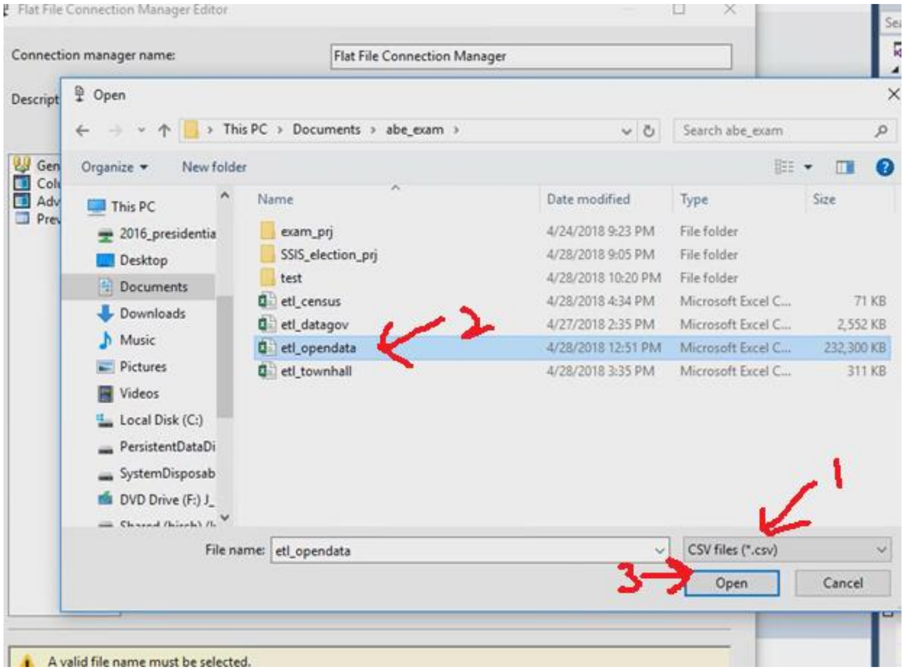
Text qualifier: <none>

Header row delimiter: {CR}{LF}

Header rows to skip: 0

☑ Column names in the first data row

⚠ Columns are not defined for this connection manager.



Flat File Connection Manager Editor

Connection manager name: Flat File Connection Manager

Open

This PC › Documents › abe_exam ›

Search abe_exam

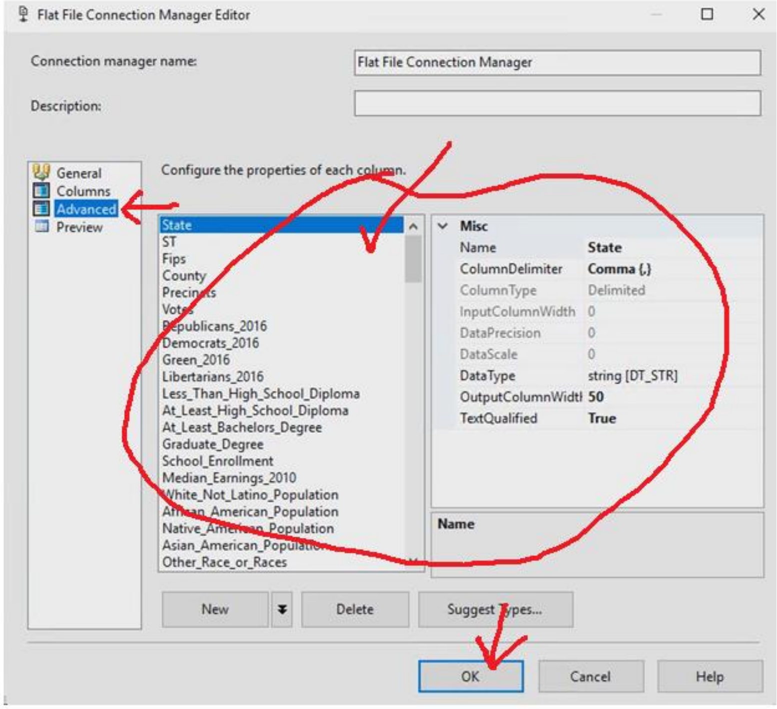
Name	Date modified	Type	Size
exam_prj	4/24/2018 9:23 PM	File folder	
SSIS_election_prj	4/28/2018 9:05 PM	File folder	
test	4/28/2018 10:20 PM	File folder	
etl_census	4/28/2018 4:34 PM	Microsoft Excel C...	71 KB
etl_datagov	4/27/2018 2:35 PM	Microsoft Excel C...	2,552 KB
etl_opendata	4/28/2018 12:51 PM	Microsoft Excel C...	232,300 KB
etl_townhall	4/28/2018 3:35 PM	Microsoft Excel C...	311 KB

File name: etl_opendata

CSV files (*.csv)

A valid file name must be selected.

Click Advanced, verify columns in my OpenDataSoft CSV file, then click OK.



Flat File Connection Manager Editor

Connection manager name: Flat File Connection Manager

Description:

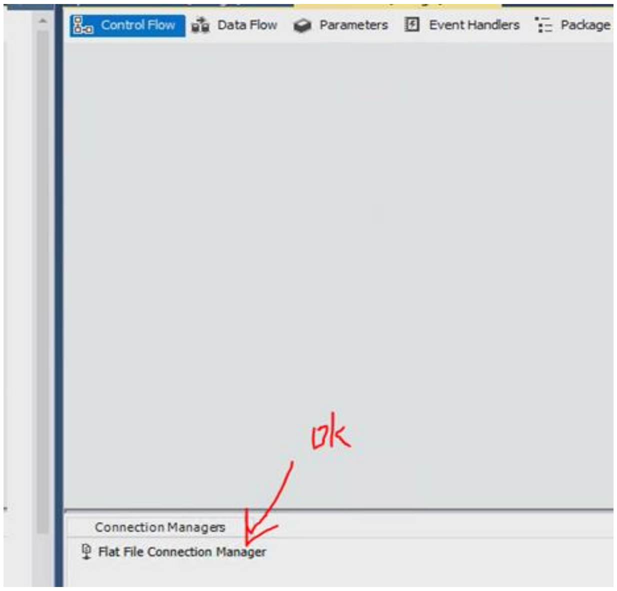
Configure the properties of each column.

State
ST
Fips
County
Precincts
Votes
Republicans_2016
Democrats_2016
Green_2016
Libertarians_2016
Less_Than_High_School_Diploma
At_Least_High_School_Diploma
At_Least_Bachelors_Degree
Graduate_Degree
School_Enrollment
Median_Earnings_2010
White_Not_Latino_Population
African_American_Population
Native_American_Population
Asian_American_Population
Other_Race_or_Races

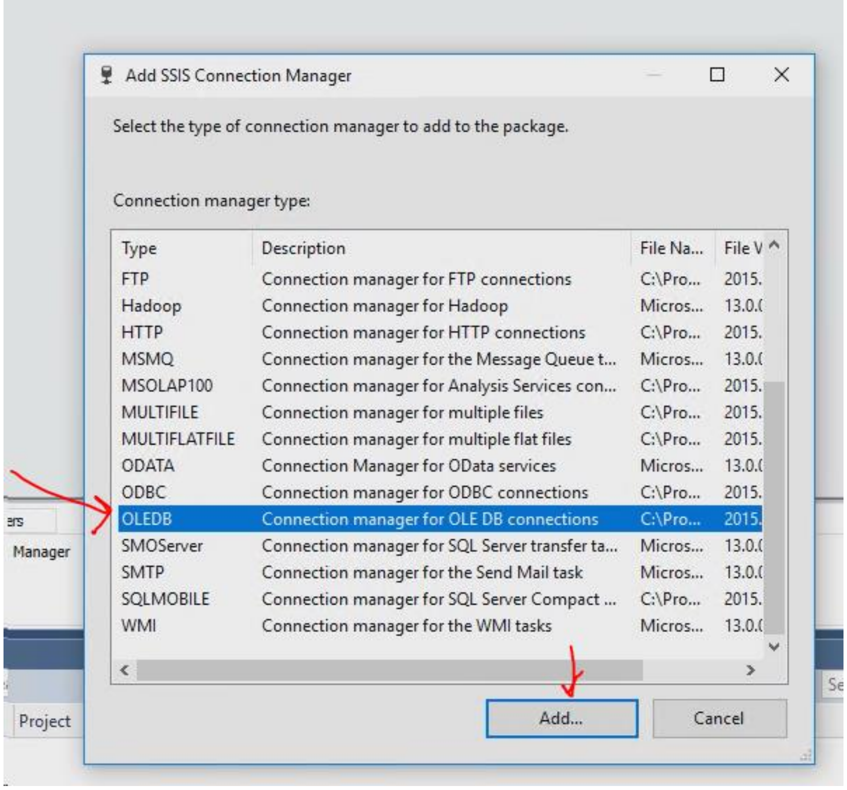
Misc	
Name	State
ColumnDelimiter	Comma {,}
ColumnType	Delimited
InputColumnWidth	0
DataPrecision	0
DataScale	0
DataType	string [DT_STR]
OutputColumnWidtl	50
TextQualified	True

Name

Next screenshot to check that Flat File Connection Manager created.



5. Repeat steps in 3, but create OLEDB connection manager.
   In Control Flow tab design surface in the middle, mouse right click, click New Connection, select OLEDB connection manager type and Add.



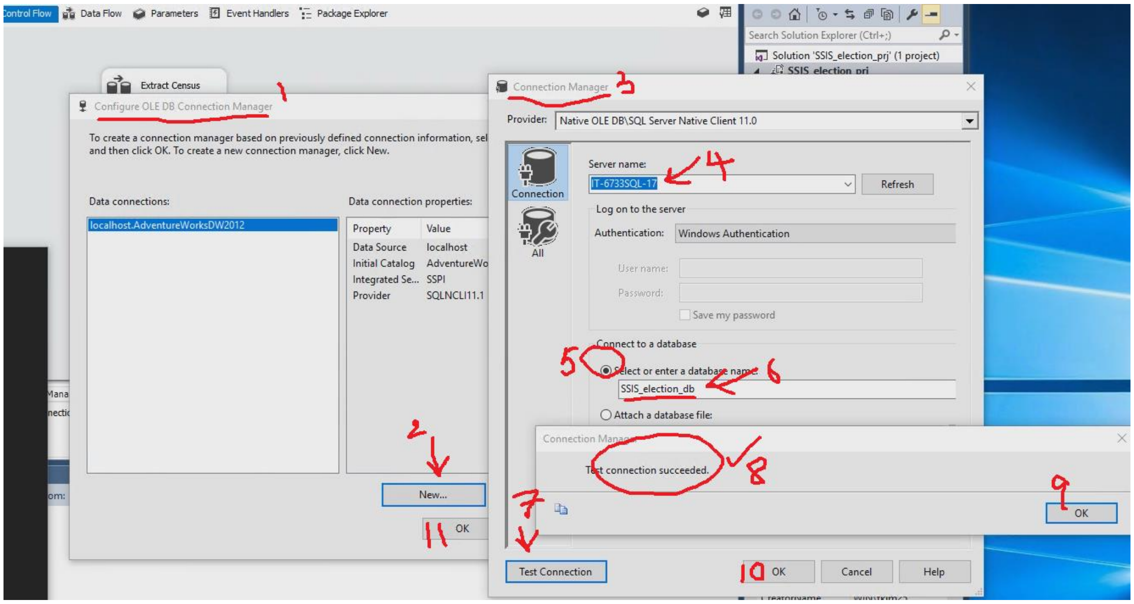
Add SSIS Connection Manager

Select the type of connection manager to add to the package.

Connection manager type:

Type	Description	File Na...	File V
FTP	Connection manager for FTP connections	C:\Pro...	2015.
Hadoop	Connection manager for Hadoop	Micros...	13.0.(
HTTP	Connection manager for HTTP connections	C:\Pro...	2015.
MSMQ	Connection manager for the Message Queue t...	Micros...	13.0.(
MSOLAP100	Connection manager for Analysis Services con...	C:\Pro...	2015.
MULTIFILE	Connection manager for multiple files	C:\Pro...	2015.
MULTIFLATFILE	Connection manager for multiple flat files	C:\Pro...	2015.
ODATA	Connection Manager for OData services	Micros...	13.0.(
ODBC	Connection manager for ODBC connections	C:\Pro...	2015.
OLEDB	Connection manager for OLE DB connections	C:\Pro...	2015.
SMOServer	Connection manager for SQL Server transfer ta...	Micros...	13.0.(
SMTP	Connection manager for the Send Mail task	Micros...	13.0.(
SQLMOBILE	Connection manager for SQL Server Compact ...	C:\Pro...	2015.
WMI	Connection manager for the WMI tasks	Micros...	13.0.(

In the next screen, create connection to my DB from step 1.



Configure OLE DB Connection Manager

To create a connection manager based on previously defined connection information, sel
and then click OK. To create a new connection manager, click New.

Data connections: localhost.AdventureWorksDW2012

Data connection properties:

Property	Value
Data Source	localhost
Initial Catalog	AdventureWo
Integrated Se...	SSPI
Provider	SQLNCLI11.1

Connection Manager

Provider: Native OLE DB\SQL Server Native Client 11.0

Server name: IT-6733SQL-17

Log on to the server

Authentication: Windows Authentication

User name:

Password:

☐ Save my password

Connect to a database

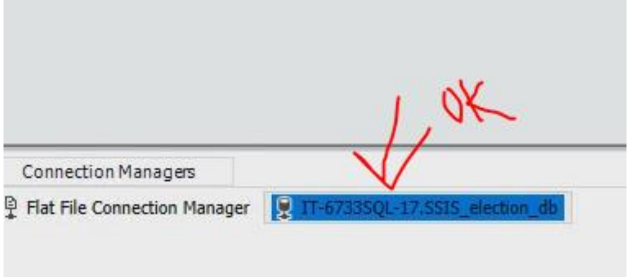
● Select or enter a database name: SSIS_election_db

○ Attach a database file:

Connection Manager

Test connection succeeded.

Next screenshot bottom of page to check job done.

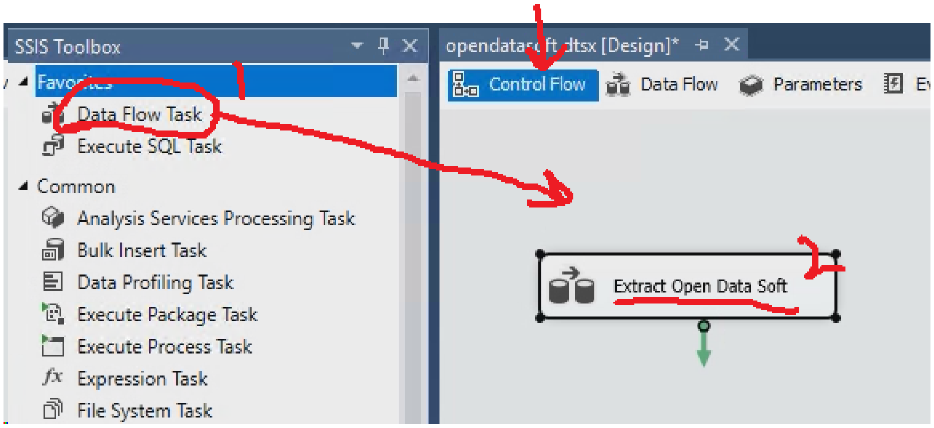


Connection Managers

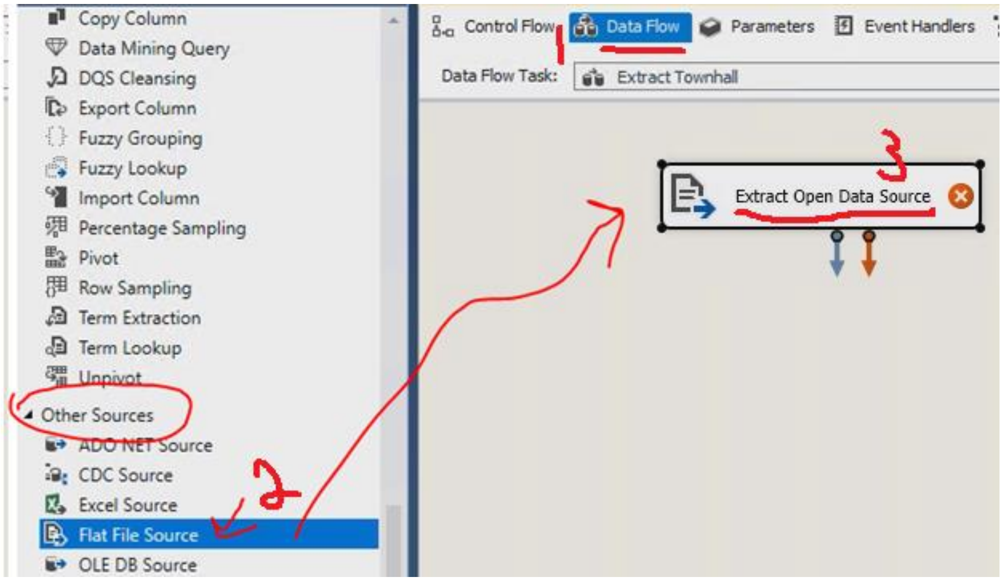
Flat File Connection Manager

IT-6733SQL-17.SSIS_election_db

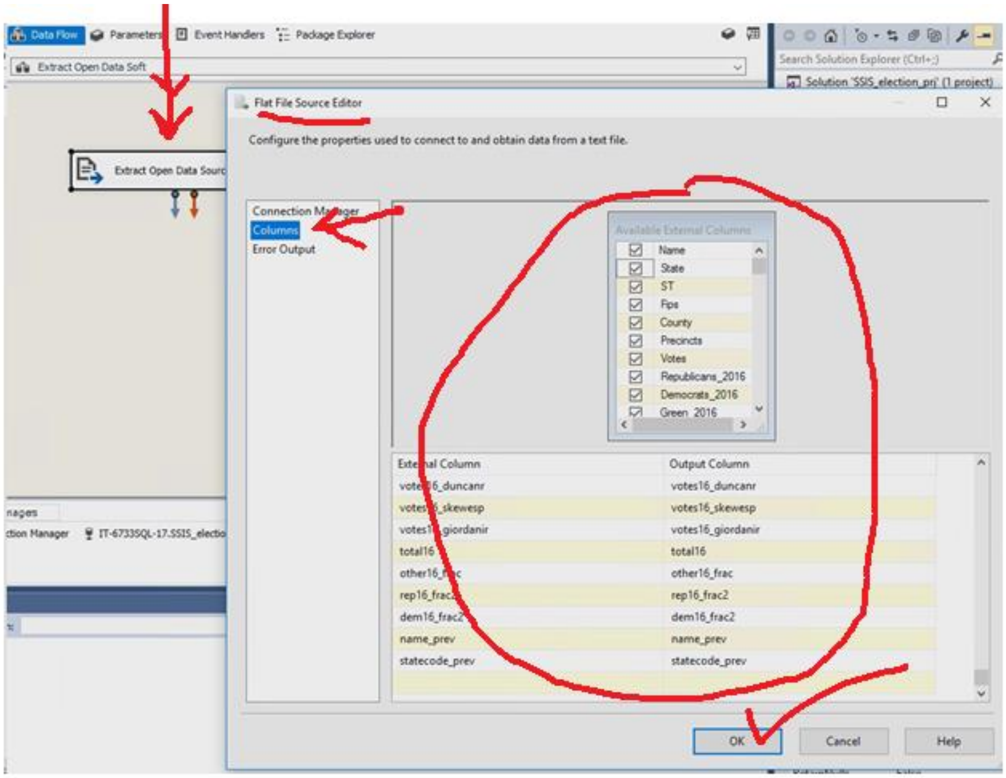
6. In Control Flow tab, click drag Data Flow Task, rename it.



7. Go to Data Flow tab, click drag Flat File source, rename it.



8. Double click Extract Open Data Source, then in next window, check columns, click OK.

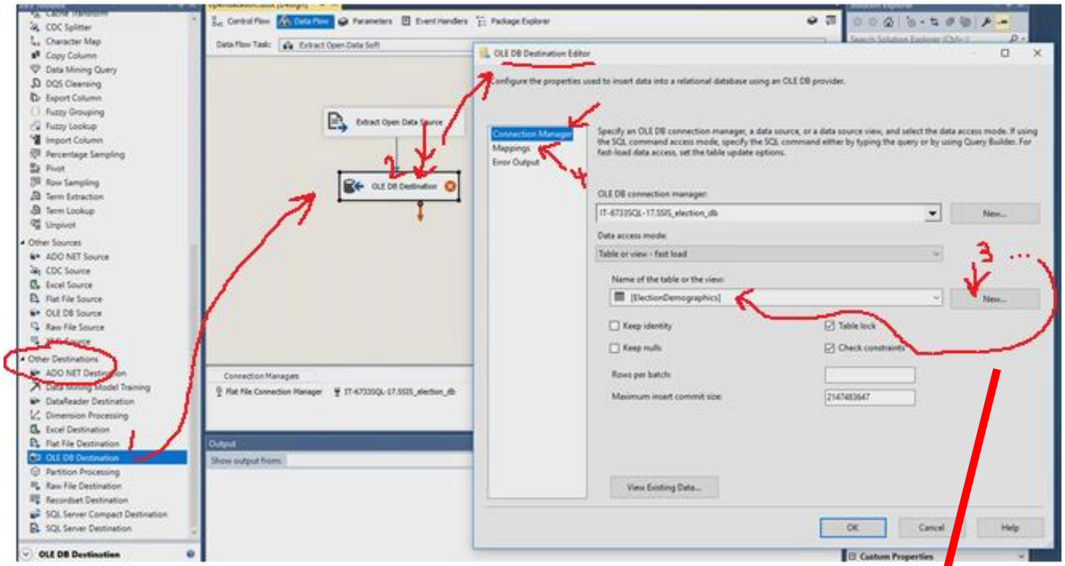


Configure the properties used to connect to and obtain data from a text file.

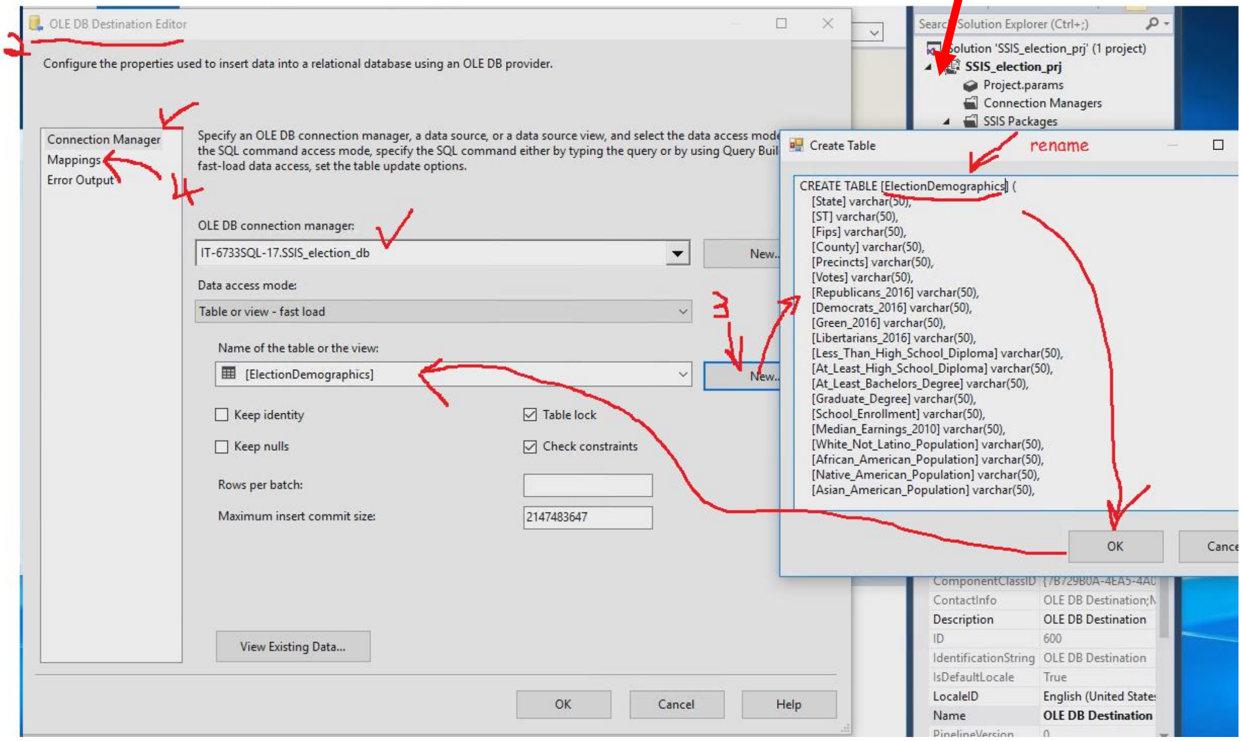
☑ Name
☑ State
☑ ST
☑ Fips
☑ County
☑ Precincts
☑ Votes
☑ Republicans_2016
☑ Democrats_2016
☑ Green_2016

External Column	Output Column
votes16_duncanr	votes16_duncanr
votes16_skewesp	votes16_skewesp
votes16_giordanir	votes16_giordanir
total16	total16
other16_frac	other16_frac
rep16_frac2	rep16_frac2
dem16_frac2	dem16_frac2
name_prev	name_prev
statecode_prev	statecode_prev

9. In same Data Flow tab, under Other Destinations, click drag OLE DB Destinations, rename it, double click it to get OLE DB Destination Editor, click Connection Manager, make sure OLE DB connection manager field shows my DB created in step 1.



10. Click New next to Name of the table of the view, rename of OLE DB Destination table to mine, click OK, then click Mappings.



Configure the properties used to insert data into a relational database using an OLE DB provider.

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

OLE DB connection manager: IT-6733SQL-17.SSIS_election_db

Data access mode: Table or view - fast load

Name of the table or the view: [ElectionDemographics]

□ Keep identity ☑ Table lock

□ Keep nulls ☑ Check constraints

Rows per batch:

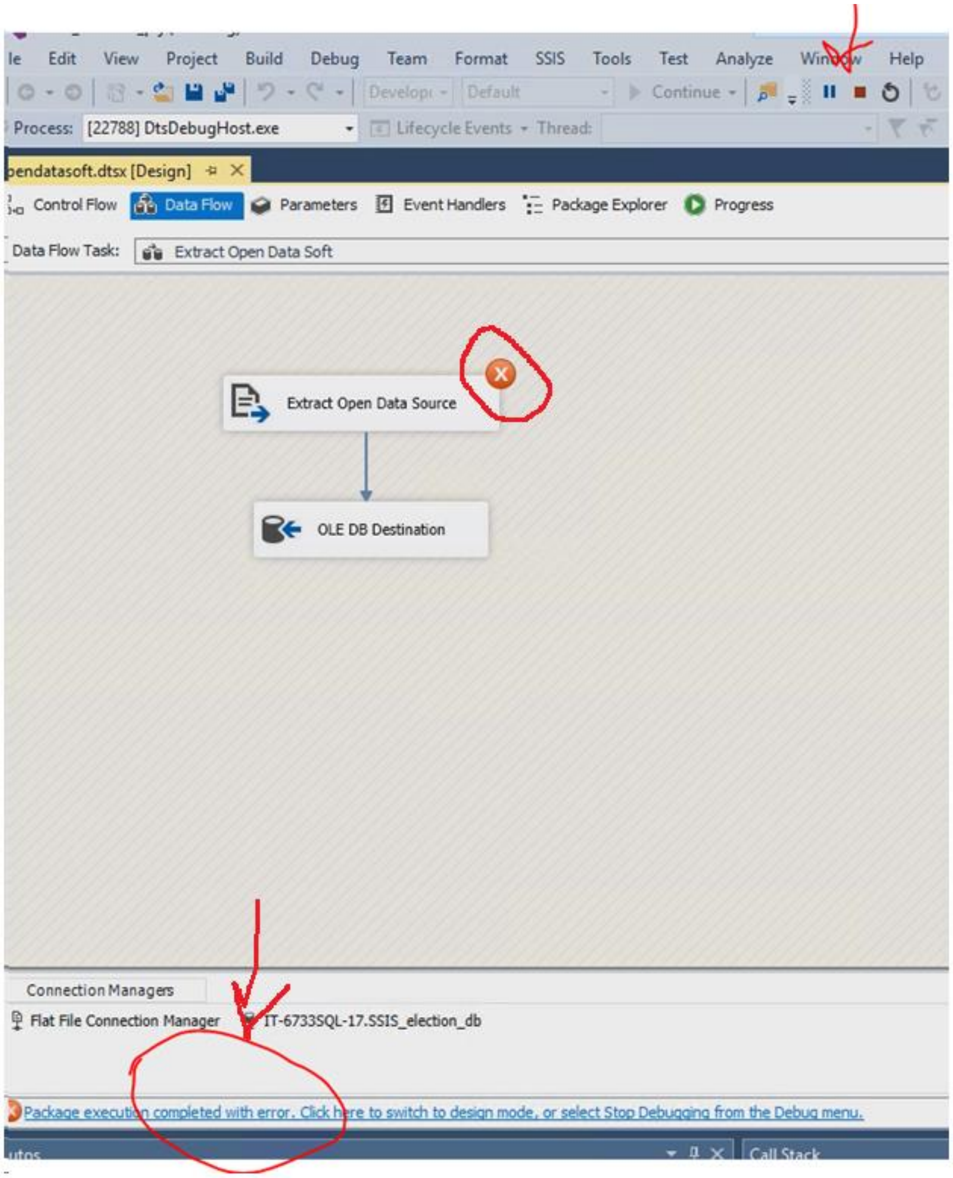
Maximum insert commit size: 2147483647

```
CREATE TABLE [ElectionDemographics] (
    [State] varchar(50),
    [ST] varchar(50),
    [Fips] varchar(50),
    [County] varchar(50),
    [Precincts] varchar(50),
    [Votes] varchar(50),
    [Republicans_2016] varchar(50),
    [Democrats_2016] varchar(50),
    [Green_2016] varchar(50),
    [Libertarians_2016] varchar(50),
    [Less_Than_High_School_Diploma] varchar(50),
    [At_Least_High_School_Diploma] varchar(50),
    [At_Least_Bachelors_Degree] varchar(50),
    [Graduate_Degree] varchar(50),
    [School_Enrollment] varchar(50),
    [Median_Earnings_2010] varchar(50),
    [White_Not_Latino_Population] varchar(50),
    [African_American_Population] varchar(50),
    [Native_American_Population] varchar(50),
    [Asian_American_Population] varchar(50),
```

ComponentClassID	{7B729B0A-4EA5-4A0
ContactInfo	OLE DB Destination;N
Description	OLE DB Destination
ID	600
IdentificationString	OLE DB Destination
IsDefaultLocale	True
LocaleID	English (United State
Name	OLE DB Destination
PipelineVersion	0

In Mappings, verify that columns of source and destination match up, click OK. (no screenshot)

11. Then click Start. But, oops, our first file load resulted in an error:


le Edit View Project Build Debug Team Format SSIS Tools Test Analyze Window Help
Develop Default Continue
Process: [22788] DtsDebugHost.exe
Lifecycle Events Thread:
pendatasoft.dtsx [Design]
Control Flow
Data Flow
Parameters
Event Handlers
Package Explorer
Progress
Data Flow Task: Extract Open Data Soft
Extract Open Data Source
OLE DB Destination
Connection Managers
Flat File Connection Manager
IT-6733SQL-17.SSIS_election_db
utos
Call Stack

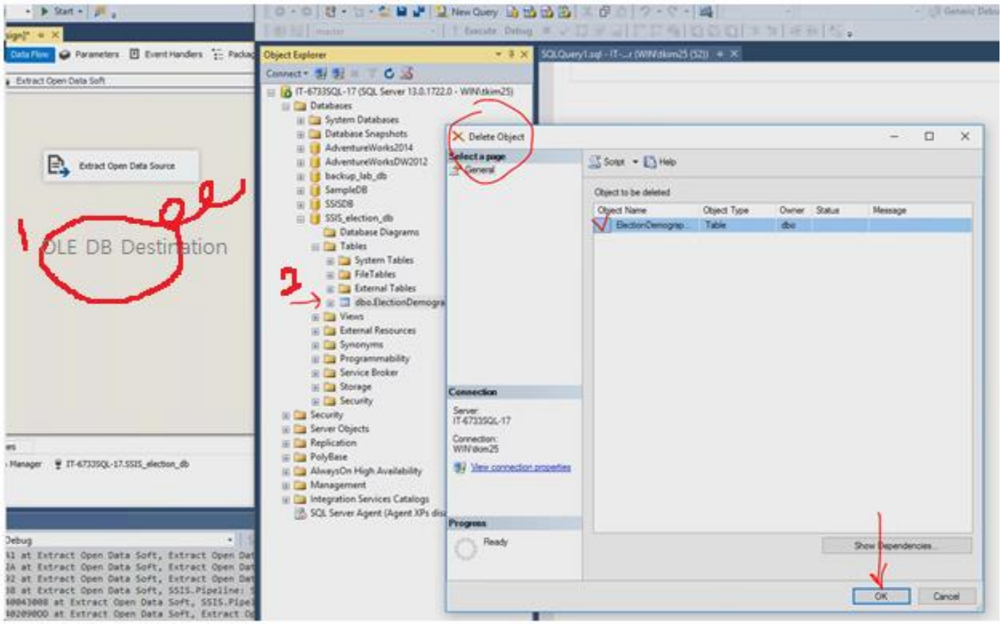
Package execution completed with error. Click here to switch to design mode, or select Stop Debugging from the Debug menu.

12. Our Error Message:

```
Error: 0xC02020A1 at Extract Open Data Soft, Extract Open Data Source [2]: Data conversion failed.
The data conversion for column "statecode_prev" returned status value 4 and status text "Text was
truncated or one or more characters had no match in the target code page.".

Error: 0xC020902A at Extract Open Data Soft, Extract Open Data Source [2]: The "Extract Open Data
Source.Outputs[Flat File Source Output].Columns[statecode_prev]" failed because truncation
occurred, and the truncation row disposition on "Extract Open Data Source.Outputs[Flat File Source
Output].Columns[statecode_prev]" specifies failure on truncation. A truncation error occurred on
the specified object of the specified component.

Error: 0xC0202092 at Extract Open Data Soft, Extract Open Data Source [2]: An error occurred while
processing file "D:\Users\tkim25\Documents\abe_exam\etl_opendata.csv" on data row 2.

Error: 0xC0047038 at Extract Open Data Soft, SSIS.Pipeline: SSIS Error Code
DTS_E_PRIMEOUTPUTFAILED.  The PrimeOutput method on Extract Open Data Source returned
error code 0xC0202092.  The component returned a failure code when the pipeline engine called
PrimeOutput(). The meaning of the failure code is defined by the component, but the error is fatal
and the pipeline stopped executing.  There may be error messages posted before this with more
information about the failure.

Information: 0x40043008 at Extract Open Data Soft, SSIS.Pipeline: Post Execute phase is beginning.

Information: 0x402090DD at Extract Open Data Soft, Extract Open Data Source [2]: The processing
of file "D:\Users\tkim25\Documents\abe_exam\etl_opendata.csv" has ended.

Information: 0x4004300B at Extract Open Data Soft, SSIS.Pipeline: "OLE DB Destination" wrote 0
rows.

Information: 0x40043009 at Extract Open Data Soft, SSIS.Pipeline: Cleanup phase is beginning.

Task failed: Extract Open Data Soft

Warning: 0x80019002 at opendatasoft: SSIS Warning Code
DTS_W_MAXIMUMERRORCOUNTREACHED.  The Execution method succeeded, but the number of
errors raised (4) reached the maximum allowed (1); resulting in failure. This occurs when the
number of errors reaches the number specified in MaximumErrorCount. Change the
MaximumErrorCount or fix the errors.

SSIS package
"D:\Users\tkim25\Documents\abe_exam\SSIS_election_prj\SSIS_election_prj\opendatasoft.dtsx"
finished: Failure.

The program '[22788] DtsDebugHost.exe: DTS' has exited with code 0 (0x0).
```

We determined that column "statecode_prev" was not necessary, so we will drop that column.

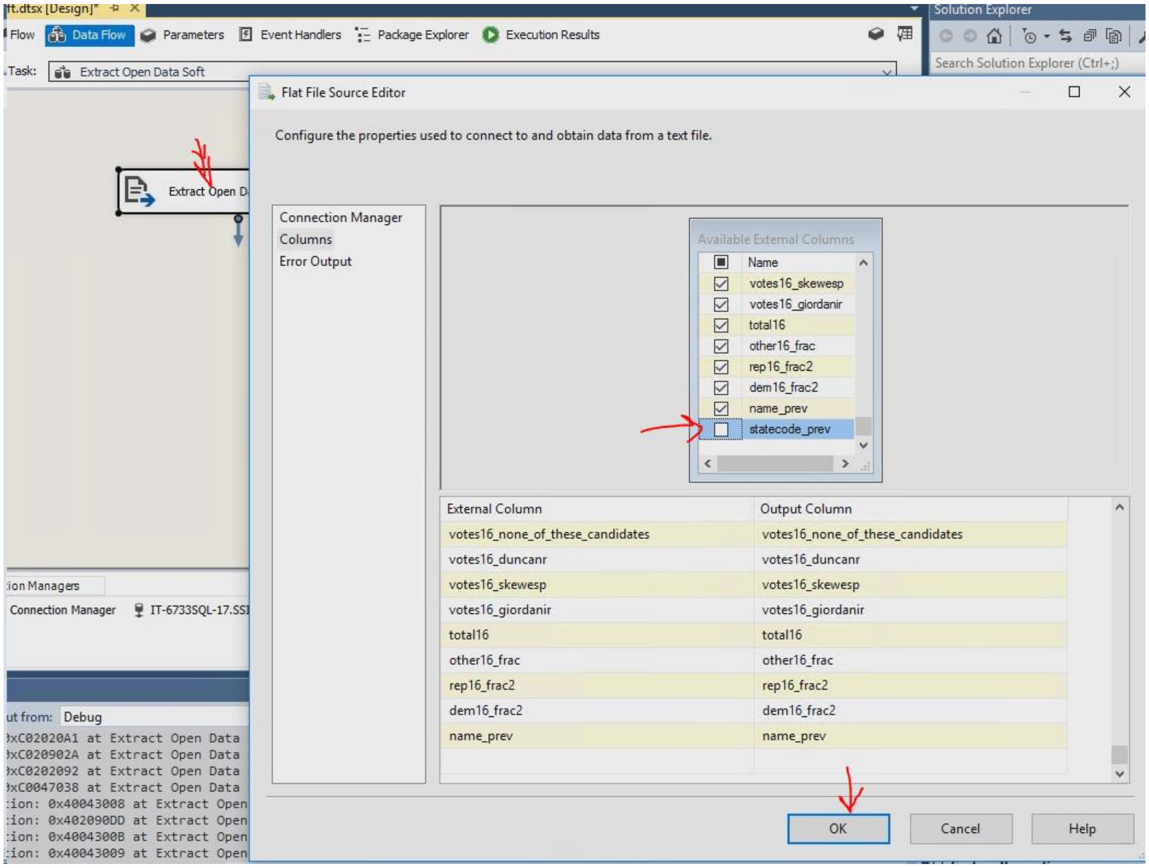
13. Delete OLE DB Destination. Go to SQL Server Management Studio, delete table dbo.ElectionDemographics.



Object to be deleted

Object Name	Object Type	Owner	Status	Message
ElectionDemograp...	Table	dbo		

14. Double click Extract Open Data Source to open Flat File Source Editor, click Columns, uncheck "statecode_prev" column, click OK.



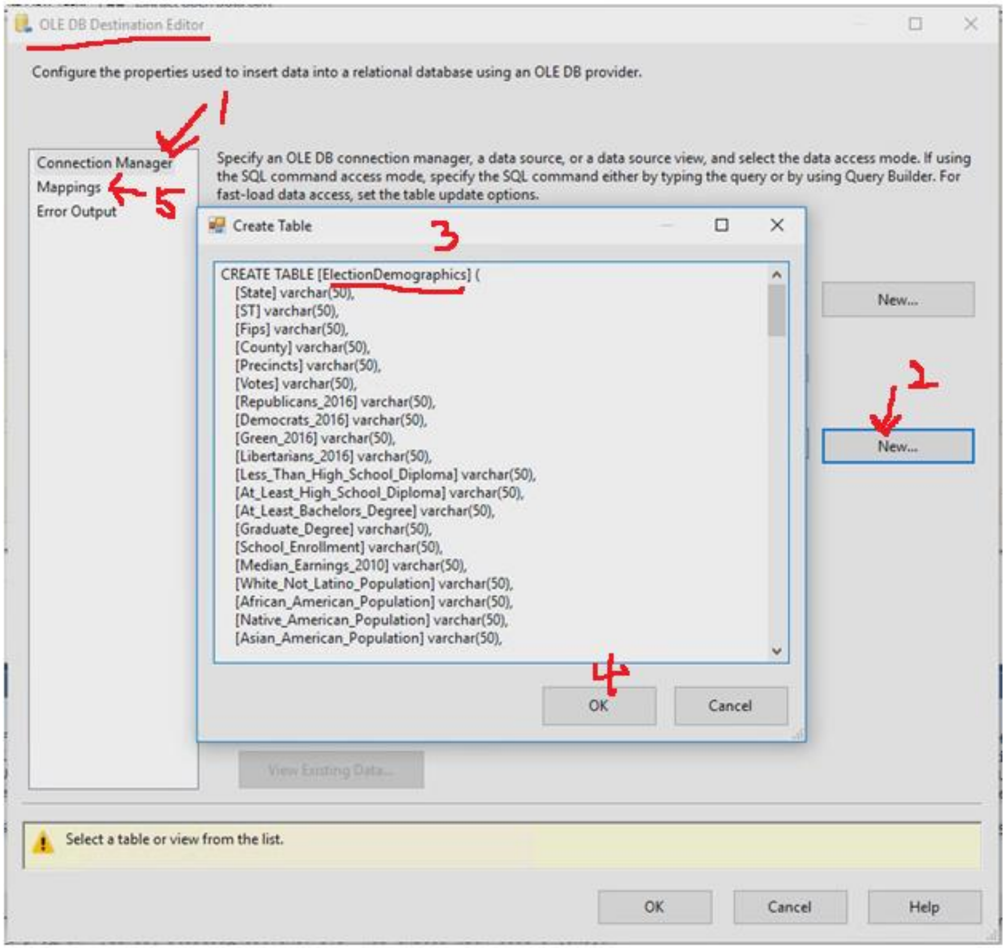
Configure the properties used to connect to and obtain data from a text file.

Available External Columns

- ■ Name
- ☑ votes16_skewesp
- ☑ votes16_giordanir
- ☑ total16
- ☑ other16_frac
- ☑ rep16_frac2
- ☑ dem16_frac2
- ☑ name_prev
- □ statecode_prev

External Column	Output Column
votes16_none_of_these_candidates	votes16_none_of_these_candidates
votes16_duncanr	votes16_duncanr
votes16_skewesp	votes16_skewesp
votes16_giordanir	votes16_giordanir
total16	total16
other16_frac	other16_frac
rep16_frac2	rep16_frac2
dem16_frac2	dem16_frac2
name_prev	name_prev

15. Re-create OLE DB Destination using step 9. In OLE DB Destination Editor window, rename destination table ("ElectionDemographics").



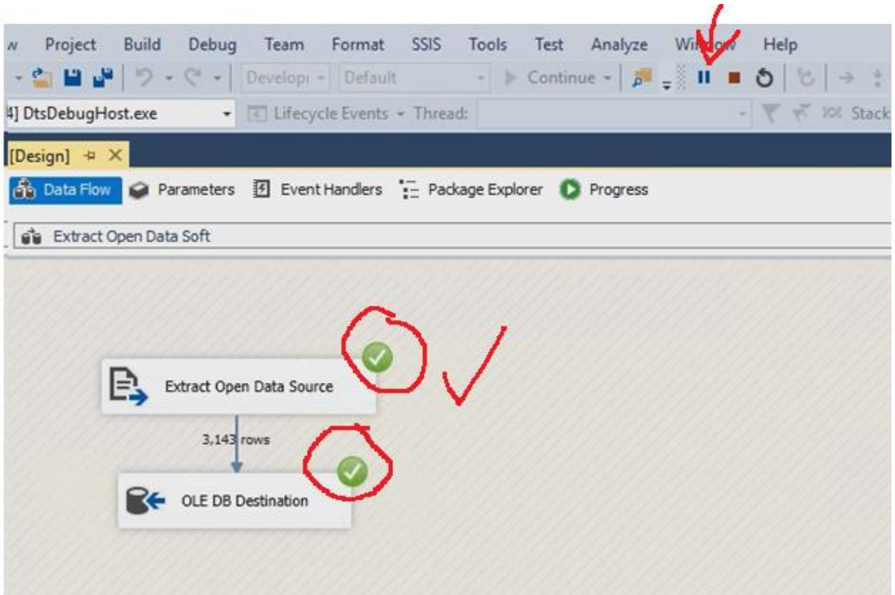
Configure the properties used to insert data into a relational database using an OLE DB provider.

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

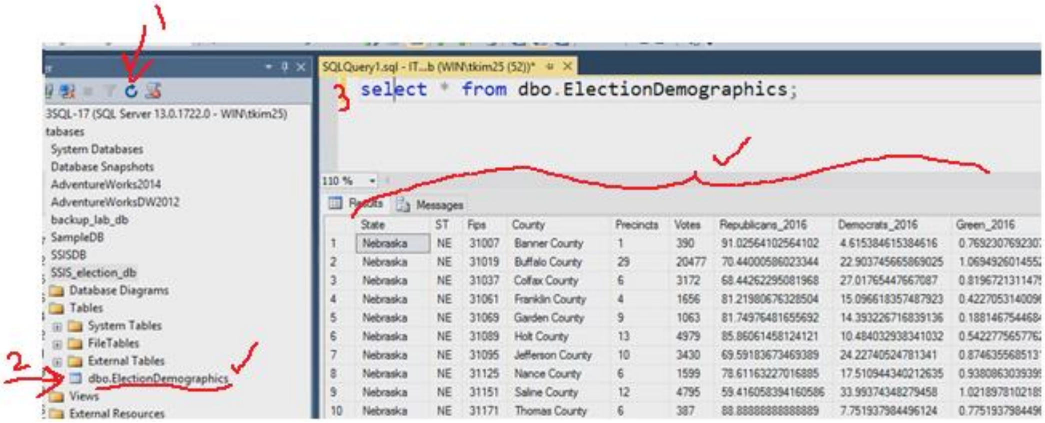
```
CREATE TABLE [ElectionDemographics] (
    [State] varchar(50),
    [ST] varchar(50),
    [Fips] varchar(50),
    [County] varchar(50),
    [Precincts] varchar(50),
    [Votes] varchar(50),
    [Republicans_2016] varchar(50),
    [Democrats_2016] varchar(50),
    [Green_2016] varchar(50),
    [Libertarians_2016] varchar(50),
    [Less_Than_High_School_Diploma] varchar(50),
    [At_Least_High_School_Diploma] varchar(50),
    [At_Least_Bachelors_Degree] varchar(50),
    [Graduate_Degree] varchar(50),
    [School_Enrollment] varchar(50),
    [Median_Earnings_2010] varchar(50),
    [White_Not_Latino_Population] varchar(50),
    [African_American_Population] varchar(50),
    [Native_American_Population] varchar(50),
    [Asian_American_Population] varchar(50),
```

Select a table or view from the list.

16. Click Start. Next screenshot to check success.



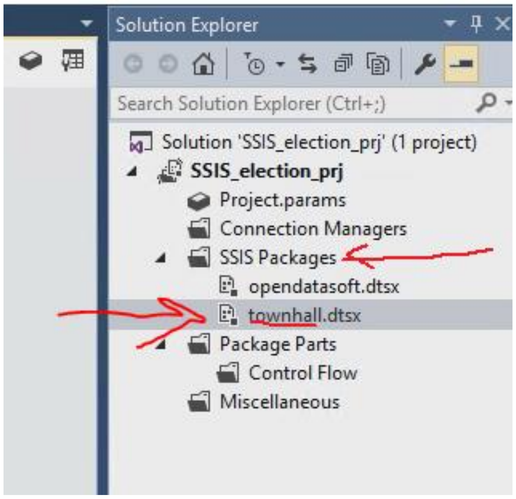
17. In SQL Server Management Studio, click refresh, see recreated table, run query, check results.



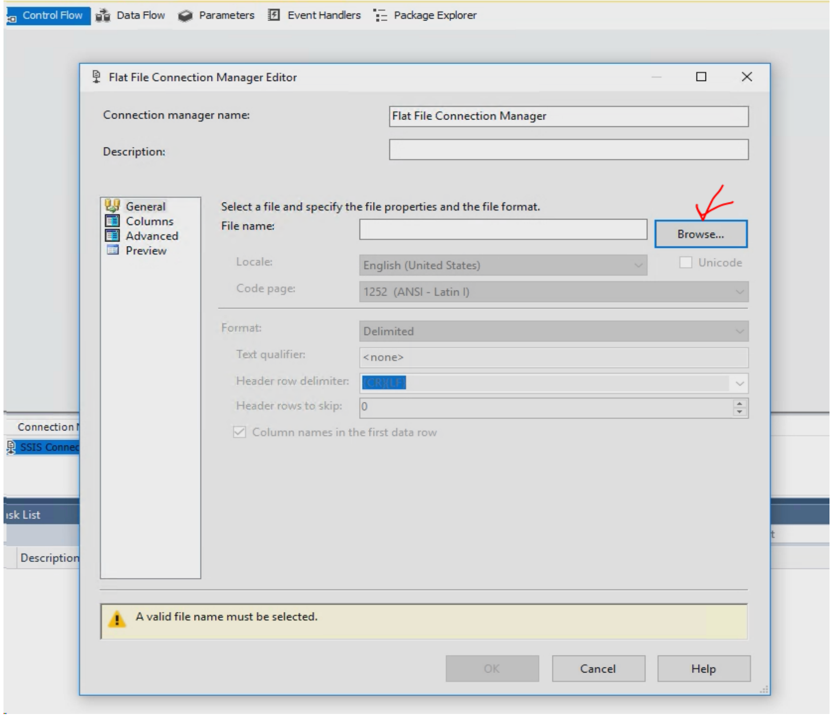
```
select * from dbo.ElectionDemographics;
```

	State	ST	Fips	County	Precincts	Votes	Republicans_2016	Democrats_2016	Green_2016
1	Nebraska	NE	31007	Banner County	1	390	91.02564102564102	4.615384615384616	0.769230769230
2	Nebraska	NE	31019	Buffalo County	29	20477	70.44000586023344	22.903745665869025	1.069492601455
3	Nebraska	NE	31037	Colfax County	6	3172	68.44262295081968	27.01765447667087	0.819672131147
4	Nebraska	NE	31061	Franklin County	4	1656	81.21980676328504	15.096618357487923	0.422705314009
5	Nebraska	NE	31069	Garden County	9	1063	81.74976481655692	14.393226716839136	0.188146754468
6	Nebraska	NE	31089	Holt County	13	4979	85.86061458124121	10.484032938341032	0.542277565776
7	Nebraska	NE	31095	Jefferson County	10	3430	69.59183673469389	24.22740524781341	0.874635568513
8	Nebraska	NE	31125	Nance County	6	1599	78.61163227016885	17.510944340212635	0.938086303939
9	Nebraska	NE	31151	Saline County	12	4795	59.416058394160586	33.99374348279458	1.021897810218
10	Nebraska	NE	31171	Thomas County	6	387	88.88888888888889	7.751937984496124	0.775193798449

18. Create a new .dtsx file for next source data: Right click SSIS Packages→ New SSIS Package → Rename it ("townhall" for this case).



19. Mouse on middle design space, right click → New Connections → select FLATFILE → double click to see next screen



Connection manager name: Flat File Connection Manager

Description:

Select a file and specify the file properties and the file format.

File name:

Locale: English (United States)

Code page: 1252 (ANSI - Latin I)

Format: Delimited

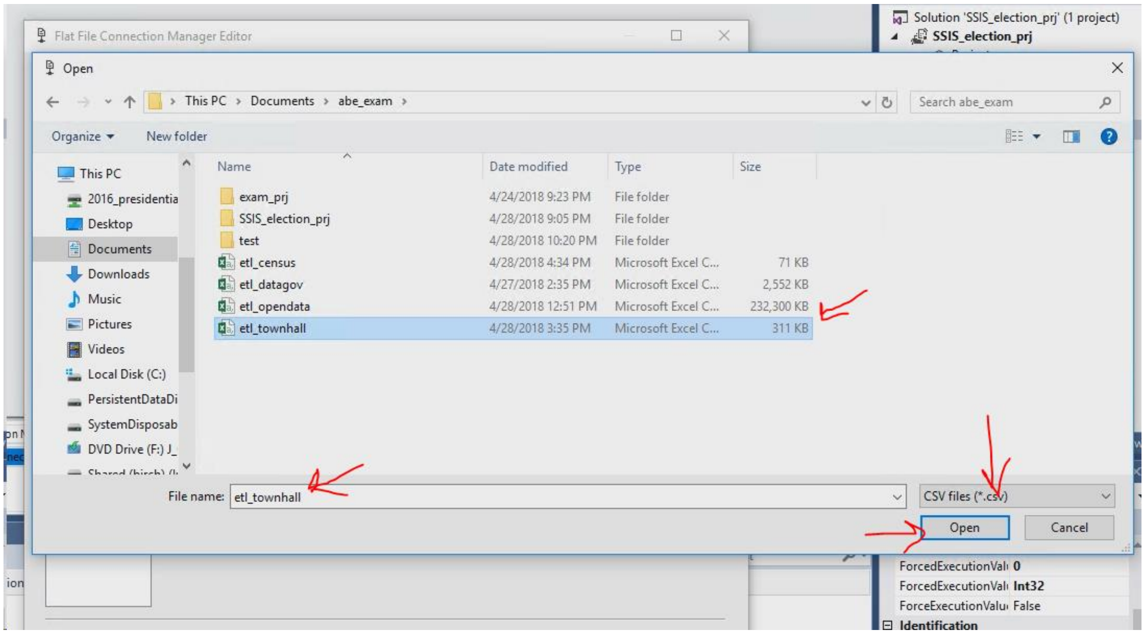
Text qualifier: <none>

Header row delimiter: {CR}{LF}

Header rows to skip: 0

☑ Column names in the first data row

A valid file name must be selected.



This PC › Documents › abe_exam ›

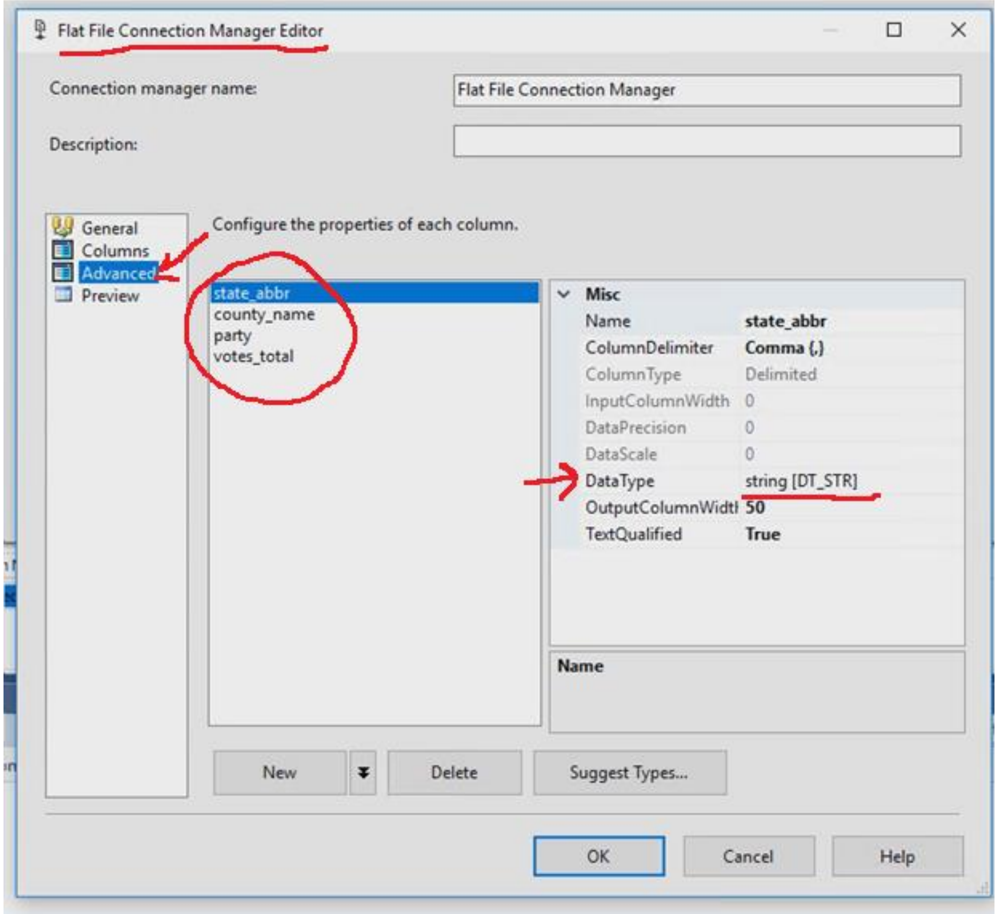
Name	Date modified	Type	Size
exam_prj	4/24/2018 9:23 PM	File folder	
SSIS_election_prj	4/28/2018 9:05 PM	File folder	
test	4/28/2018 10:20 PM	File folder	
etl_census	4/28/2018 4:34 PM	Microsoft Excel C...	71 KB
etl_datagov	4/27/2018 2:35 PM	Microsoft Excel C...	2,552 KB
etl_opendata	4/28/2018 12:51 PM	Microsoft Excel C...	232,300 KB
etl_townhall	4/28/2018 3:35 PM	Microsoft Excel C...	311 KB

File name: etl_townhall

CSV files (*.csv)

ForcedExecutionVal 0
ForcedExecutionVal Int32
ForceExecutionValu False

20. Click Advanced, verify columns in my Townhall CSV file, click OK.



Flat File Connection Manager Editor

Connection manager name: Flat File Connection Manager

Description:

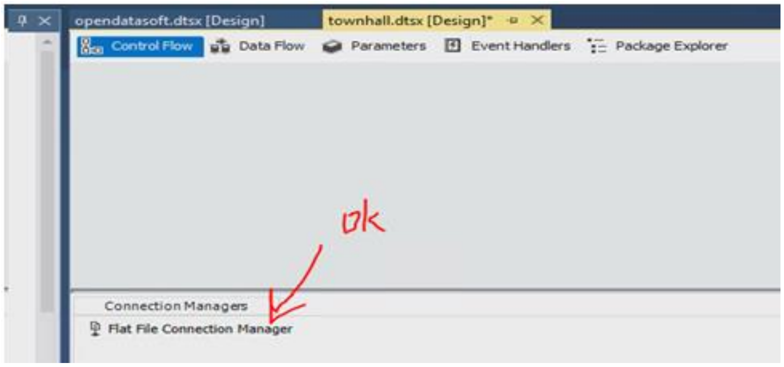
Configure the properties of each column.

state_abbr
county_name
party
votes_total

Misc	
Name	state_abbr
ColumnDelimiter	Comma {,}
ColumnType	Delimited
InputColumnWidth	0
DataPrecision	0
DataScale	0
DataType	string [DT_STR]
OutputColumnWidtl	50
TextQualified	True

Name

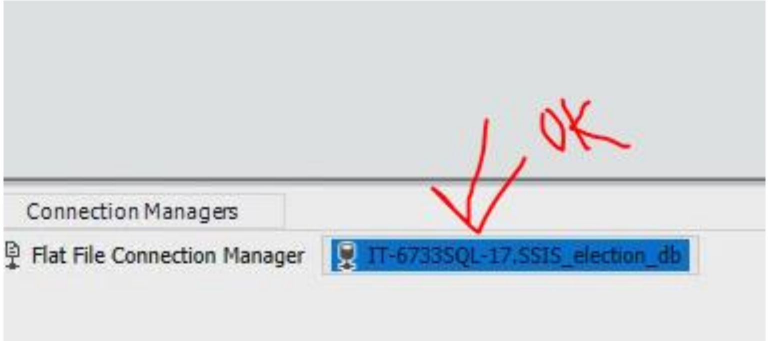
Next screen to check Flat File Conncection Manager for townhall.dtsx has been created.



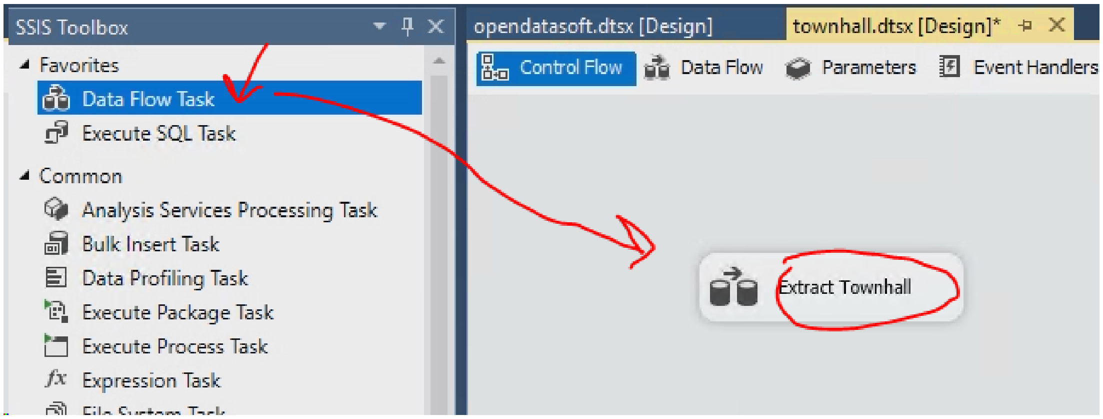
Connection Managers

Flat File Connection Manager

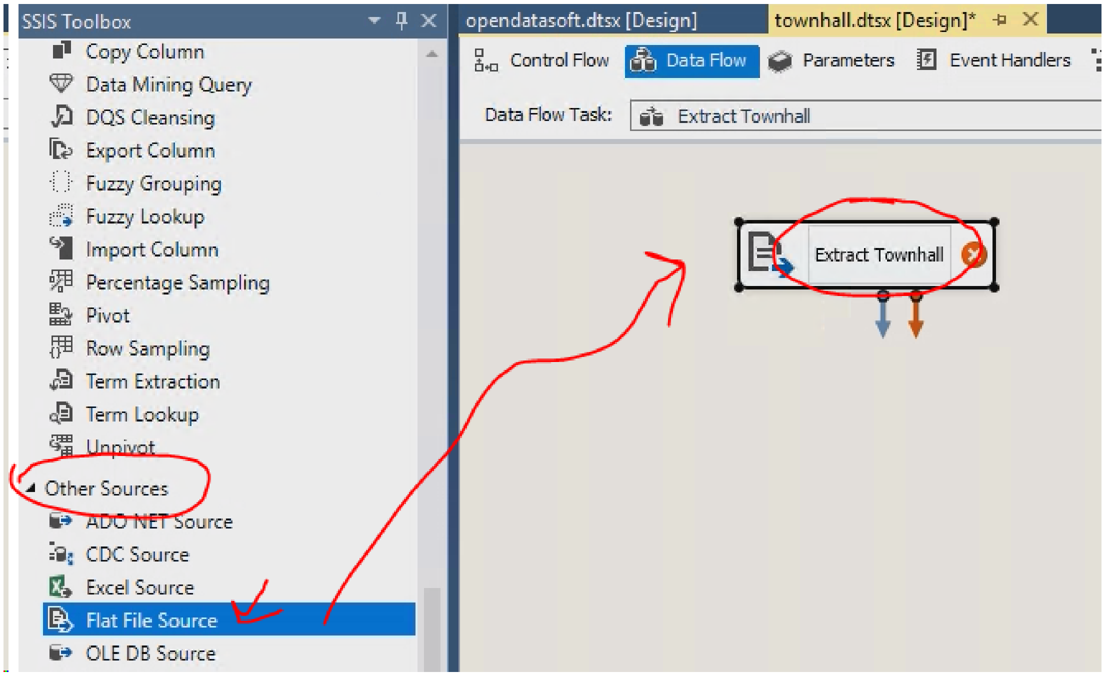
21. Repeat step 5 for OLE DB Connection Manager creation. Next screen to confirm creation.

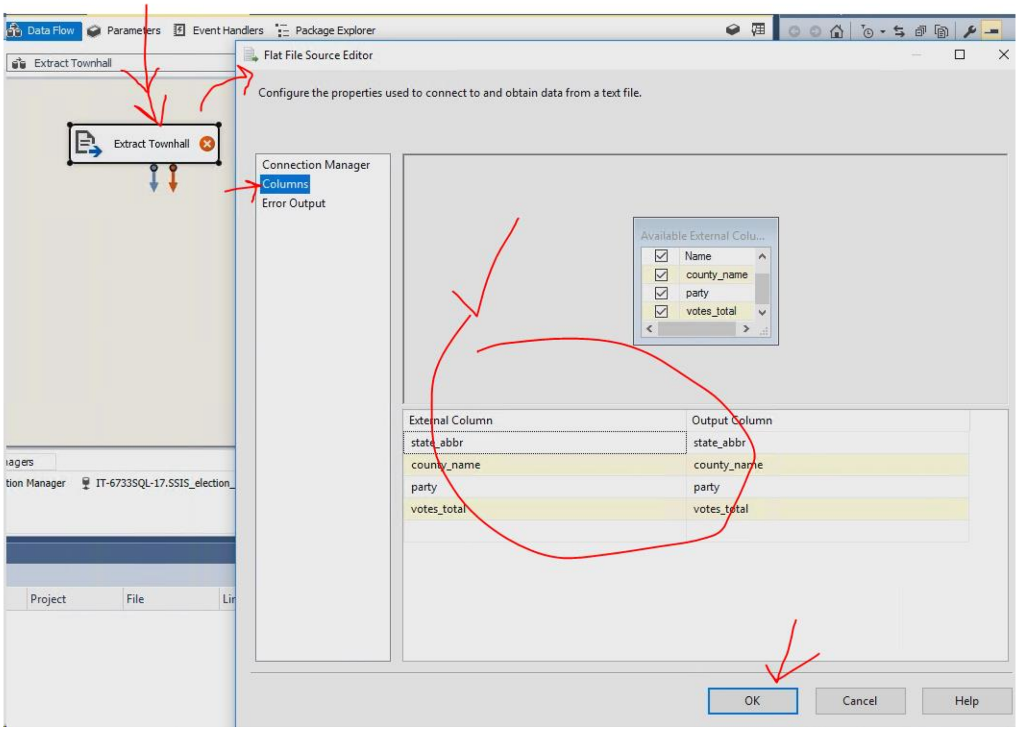


22. Repeat step to create Data Flow Task for Townhall.



23. Repeat step to create Flat File Source for Townhall.





Flat File Source Editor

Configure the properties used to connect to and obtain data from a text file.

Connection Manager
Columns
Error Output

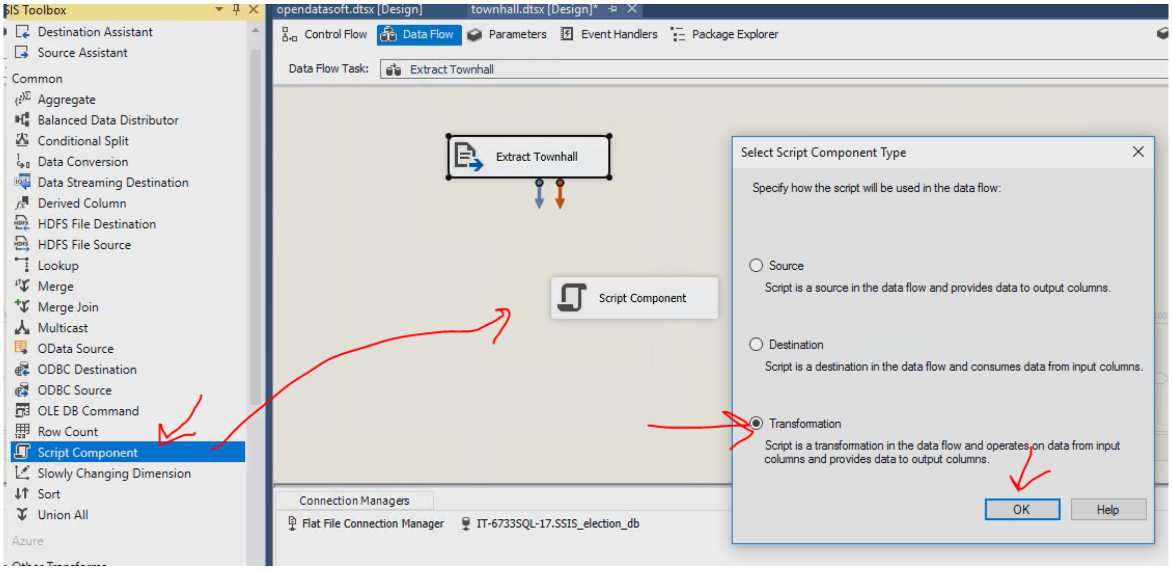
Available External Colu...

☑ Name
☑ county_name
☑ party
☑ votes_total

External Column	Output Column
state_abbr	state_abbr
county_name	county_name
party	party
votes_total	votes_total

OK   Cancel   Help

24. Click drag Script Component to add another column (state_abbr + county_name) under column name “combined” to townhall destination table.



Select Script Component Type

Specify how the script will be used in the data flow:

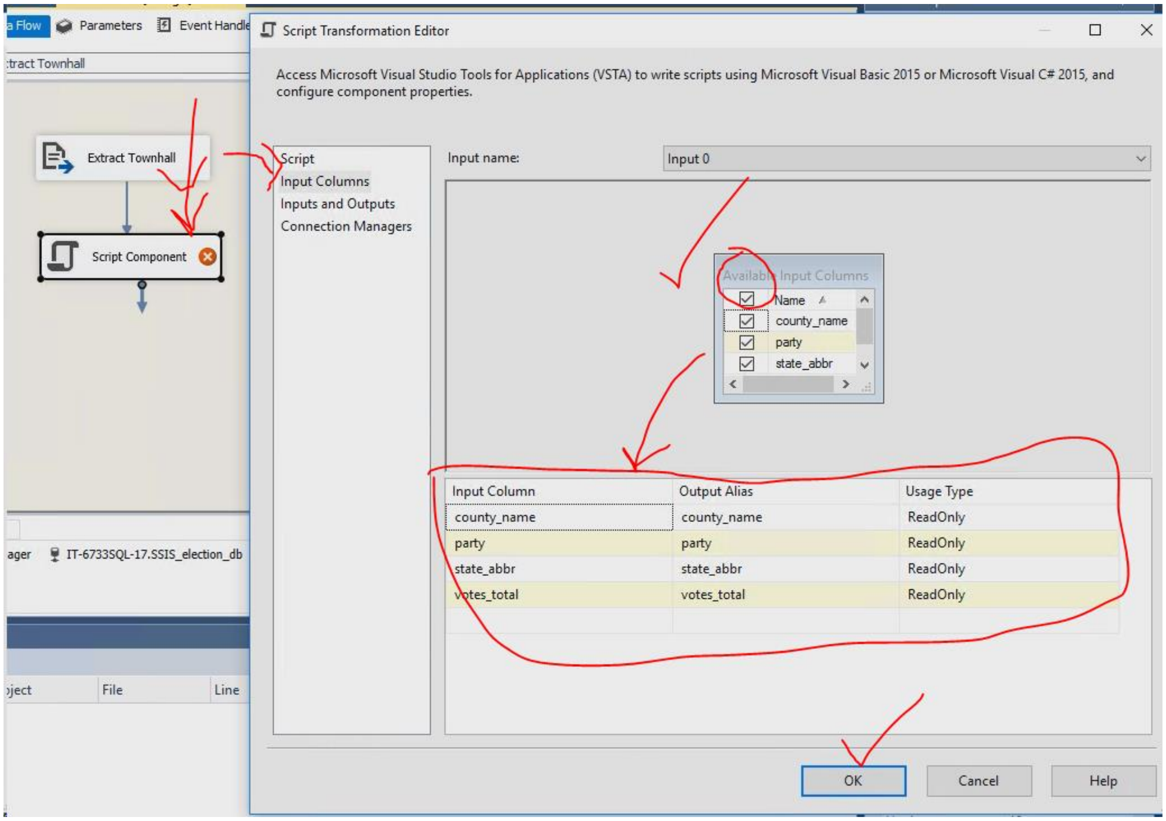
○ Source
Script is a source in the data flow and provides data to output columns.

○ Destination
Script is a destination in the data flow and consumes data from input columns.

● Transformation
Script is a transformation in the data flow and operates on data from input columns and provides data to output columns.

OK   Help

25. Double click Script Component, click Input Columns, check Name, review Input Column and Output Column, click OK.



Script Transformation Editor

Access Microsoft Visual Studio Tools for Applications (VSTA) to write scripts using Microsoft Visual Basic 2015 or Microsoft Visual C# 2015, and configure component properties.

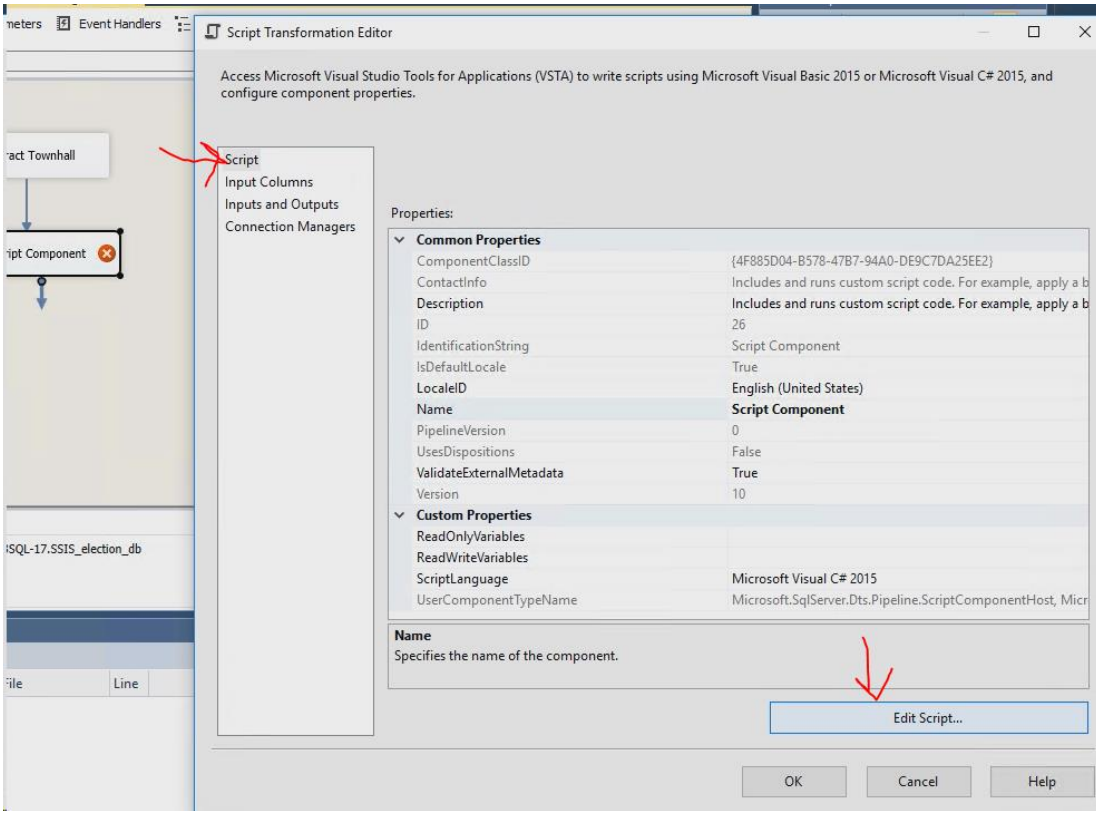
Input name: Input 0

Available Input Columns

	Name
☑	county_name
☑	party
☑	state_abbr

Input Column	Output Alias	Usage Type
county_name	county_name	ReadOnly
party	party	ReadOnly
state_abbr	state_abbr	ReadOnly
votes_total	votes_total	ReadOnly

26. Click Script, Edit Script.



Script Transformation Editor

Access Microsoft Visual Studio Tools for Applications (VSTA) to write scripts using Microsoft Visual Basic 2015 or Microsoft Visual C# 2015, and configure component properties.

Properties:

Common Properties	
ComponentClassID	{4F885D04-B578-47B7-94A0-DE9C7DA25EE2}
ContactInfo	Includes and runs custom script code. For example, apply a b
Description	Includes and runs custom script code. For example, apply a b
ID	26
IdentificationString	Script Component
IsDefaultLocale	True
LocaleID	English (United States)
Name	Script Component
PipelineVersion	0
UsesDispositions	False
ValidateExternalMetadata	True
Version	10
Custom Properties	
ReadOnlyVariables	
ReadWriteVariables	
ScriptLanguage	Microsoft Visual C# 2015
UserComponentTypeName	Microsoft.SqlServer.Dts.Pipeline.ScriptComponentHost, Micr

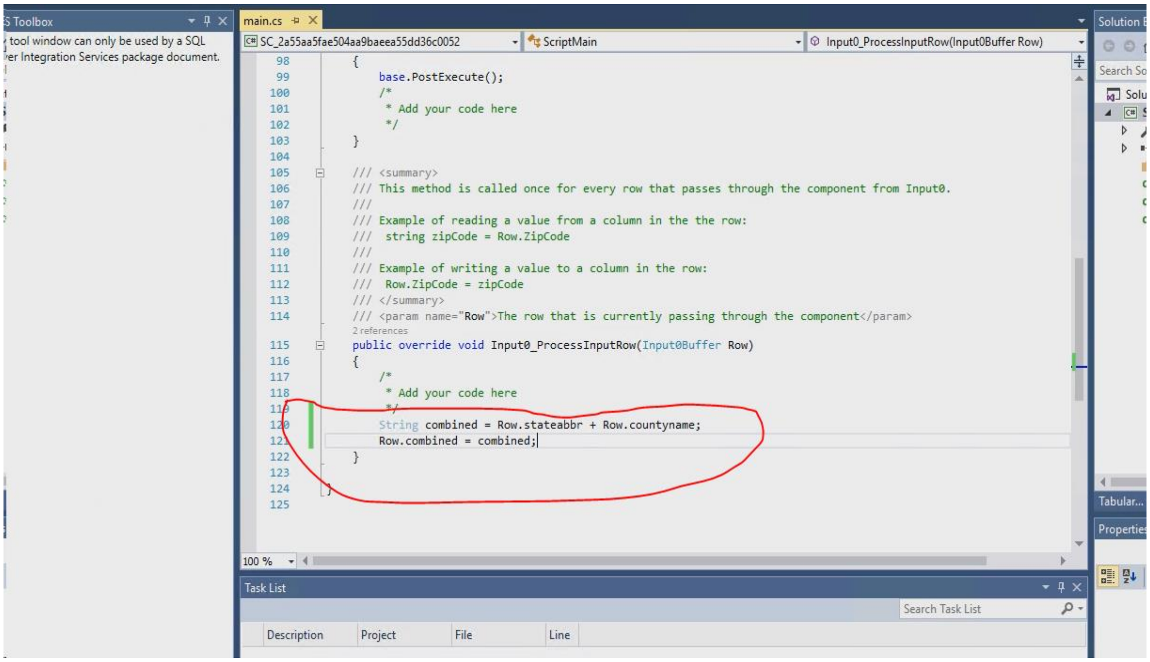
Name
Specifies the name of the component.

27. Then new window opens, under comment /* Add your code here*/ add my code:

    String combined = Row.stateabbr + Row.countyname;

    Row.combined = combined;

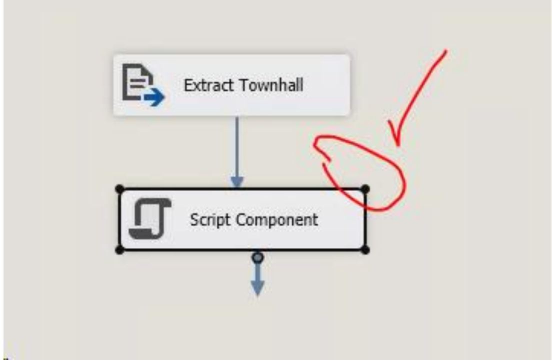
    Then save.



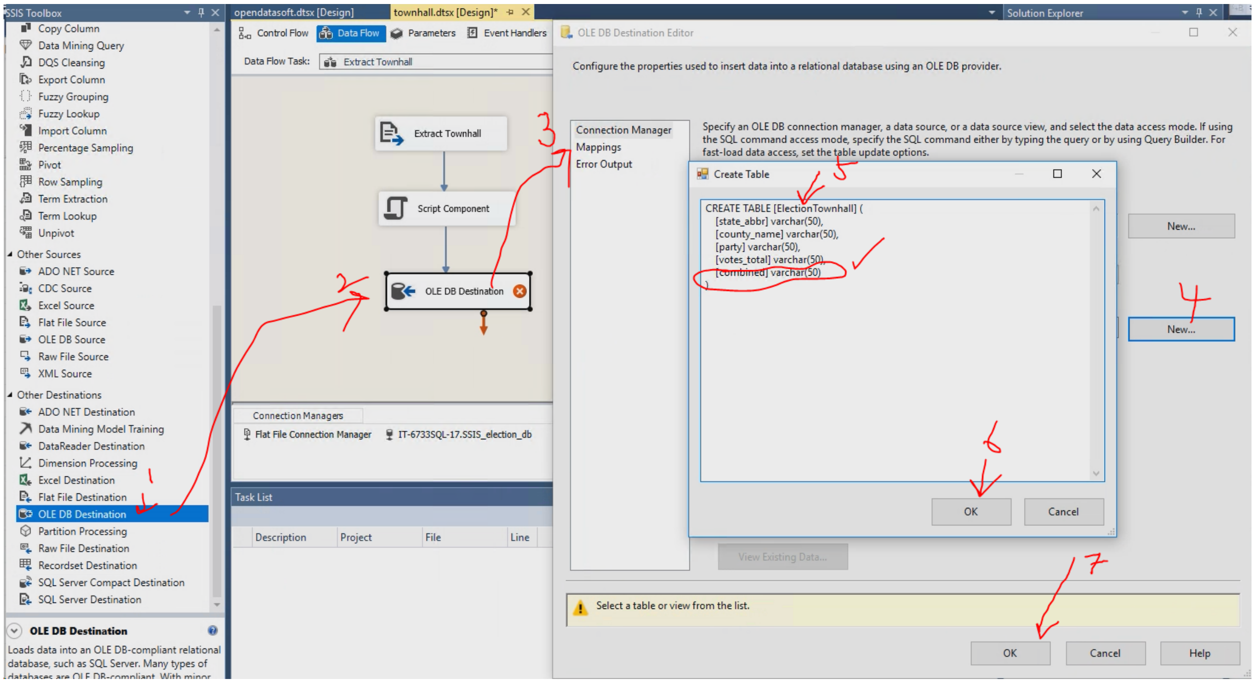
tool window can only be used by a SQL
ver Integration Services package document.

```
        {
            base.PostExecute();
            /*
             * Add your code here
             */
        }

        /// <summary>
        /// This method is called once for every row that passes through the component from Input0.
        ///
        /// Example of reading a value from a column in the the row:
        ///  string zipCode = Row.ZipCode
        ///
        /// Example of writing a value to a column in the row:
        ///  Row.ZipCode = zipCode
        /// </summary>
        /// <param name="Row">The row that is currently passing through the component</param>
        2 references
        public override void Input0_ProcessInputRow(Input0Buffer Row)
        {
            /*
             * Add your code here
             */
            String combined = Row.stateabbr + Row.countyname;
            Row.combined = combined;
        }

}
```

28. Then click OK on previous Script Transformation Editor pop up. Check that red X disappears from Script Component on design space.



29. Create OLE DB Destination, click New to rename destination table to “ElectionTownhall” and check that newly added column shows, then OK, OK.



OLE DB Destination

Loads data into an OLE DB-compliant relational database, such as SQL Server. Many types of databases are OLE DB-compliant. With minor

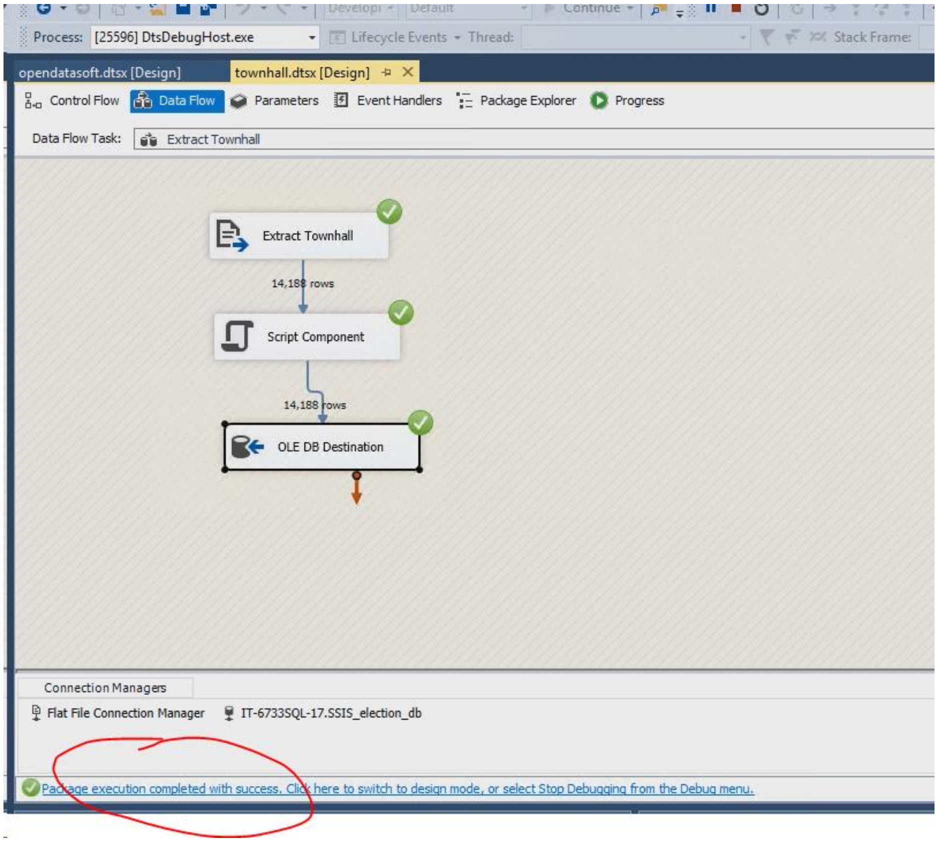
Configure the properties used to insert data into a relational database using an OLE DB provider.

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

```
CREATE TABLE [ElectionTownhall] (
    [state_abbr] varchar(50),
    [county_name] varchar(50),
    [party] varchar(50),
    [votes_total] varchar(50),
    [combined] varchar(50)
)
```

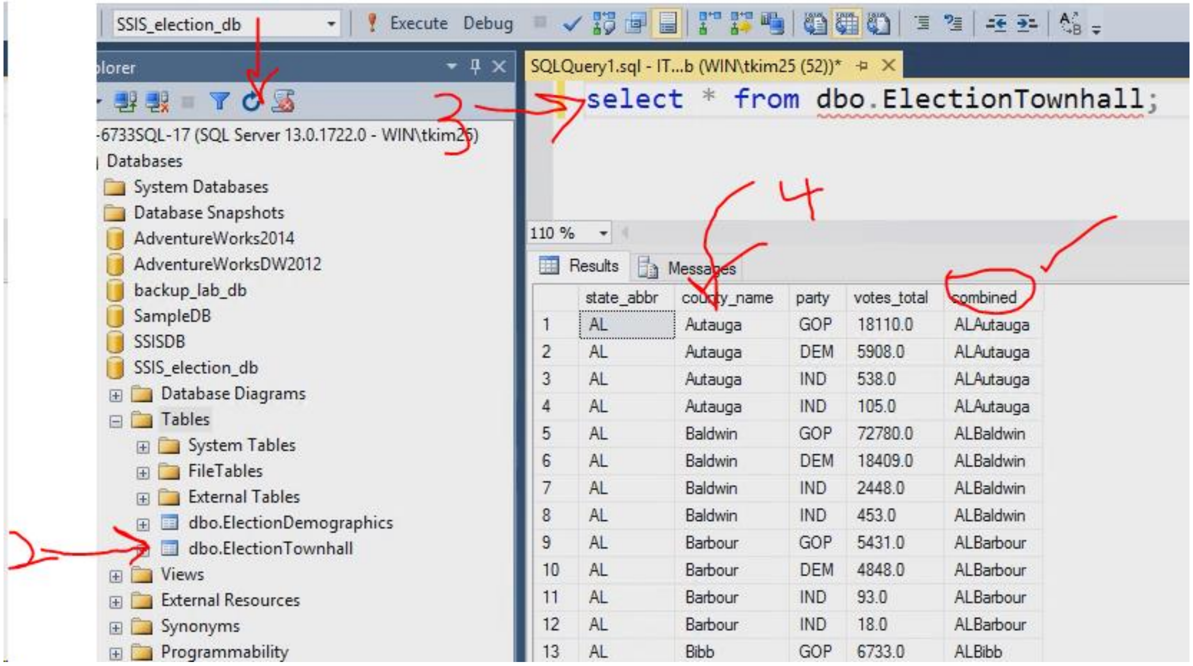
Select a table or view from the list.

30. Click Start then check successful execution.



Package execution completed with success. Click here to switch to design mode, or select Stop Debugging from the Debug menu.

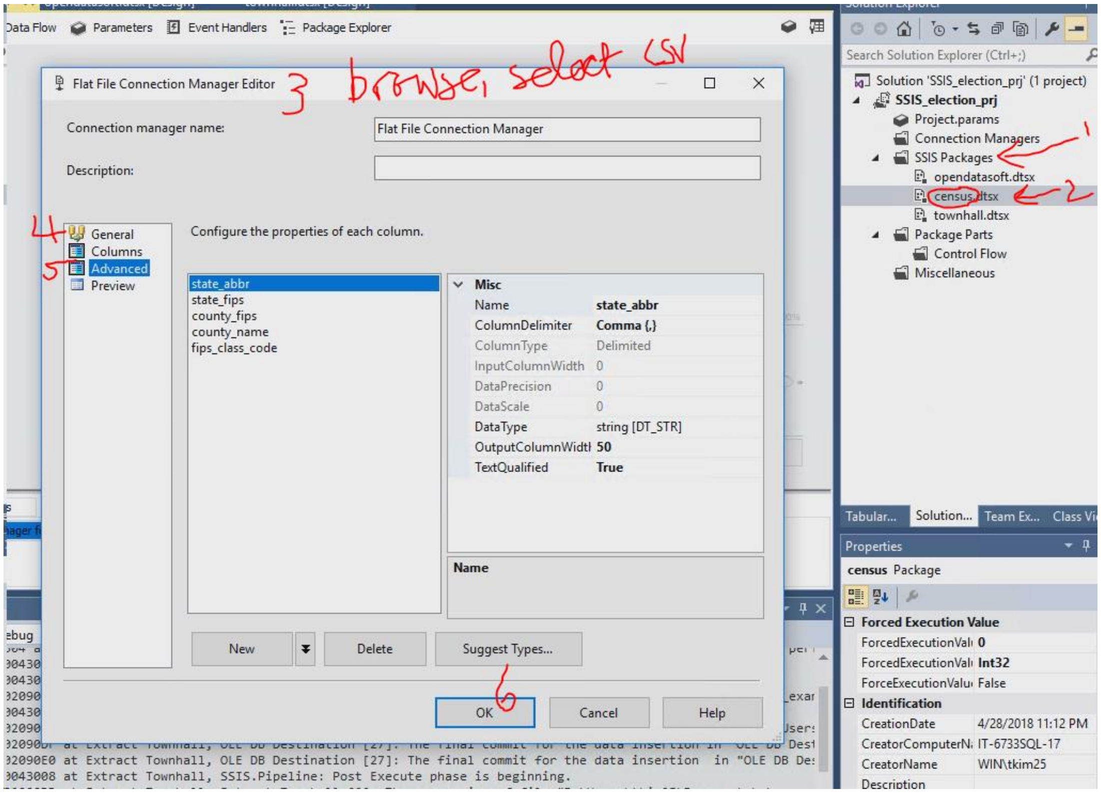
31. Go to SQL Server Management Studio to check that new destination table is there and run test query to verify newly added column appears.



```
select * from dbo.ElectionTownhall;
```

	state_abbr	county_name	party	votes_total	combined
1	AL	Autauga	GOP	18110.0	ALAutauga
2	AL	Autauga	DEM	5908.0	ALAutauga
3	AL	Autauga	IND	538.0	ALAutauga
4	AL	Autauga	IND	105.0	ALAutauga
5	AL	Baldwin	GOP	72780.0	ALBaldwin
6	AL	Baldwin	DEM	18409.0	ALBaldwin
7	AL	Baldwin	IND	2448.0	ALBaldwin
8	AL	Baldwin	IND	453.0	ALBaldwin
9	AL	Barbour	GOP	5431.0	ALBarbour
10	AL	Barbour	DEM	4848.0	ALBarbour
11	AL	Barbour	IND	93.0	ALBarbour
12	AL	Barbour	IND	18.0	ALBarbour
13	AL	Bibb	GOP	6733.0	ALBibb

32. SSIS Packages → New SSIS Package → Rename as census → right click design space → select Flat File Connection Manager



Connection manager name: Flat File Connection Manager

Description:

Configure the properties of each column.

state_abbr
state_fips
county_fips
county_name
fips_class_code

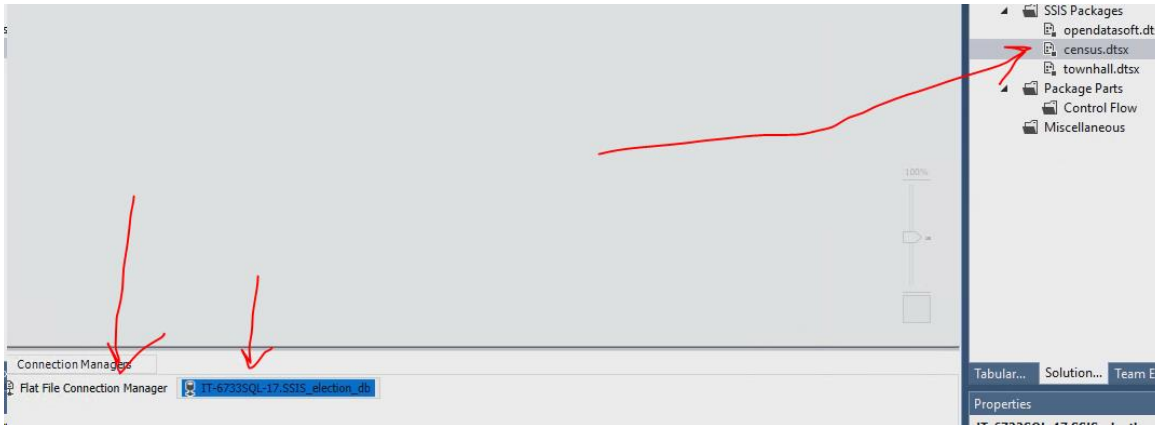
Misc	
Name	state_abbr
ColumnDelimiter	Comma {,}
ColumnType	Delimited
InputColumnWidth	0
DataPrecision	0
DataScale	0
DataType	string [DT_STR]
OutputColumnWidt	50
TextQualified	True

Name

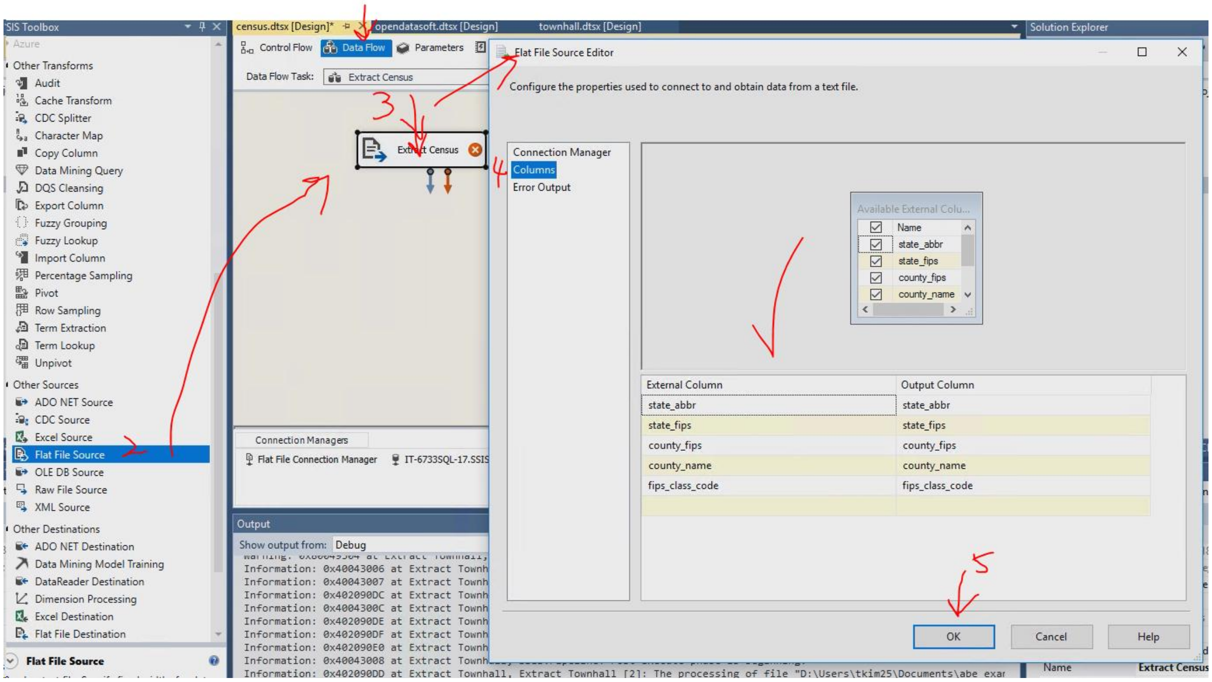
Forced Execution Value	
ForcedExecutionVal	0
ForcedExecutionVal	Int32
ForceExecutionValu	False
Identification	
CreationDate	4/28/2018 11:12 PM
CreatorComputerN	IT-6733SQL-17
CreatorName	WIN\tkim25
Description	

```
32090E0 at Extract Townhall, OLE DB Destination [27]: The final commit for the data insertion  in "OLE DB De:
3043008 at Extract Townhall, SSIS.Pipeline: Post Execute phase is beginning.
```

33. Repeat above steps to create Flat File Connection Manager, OLE DB Connection Manager, Data Flow Task, Flat File.



34. After creating a Data Flow Task in Control Flow tab, create Flat File Source for Census.

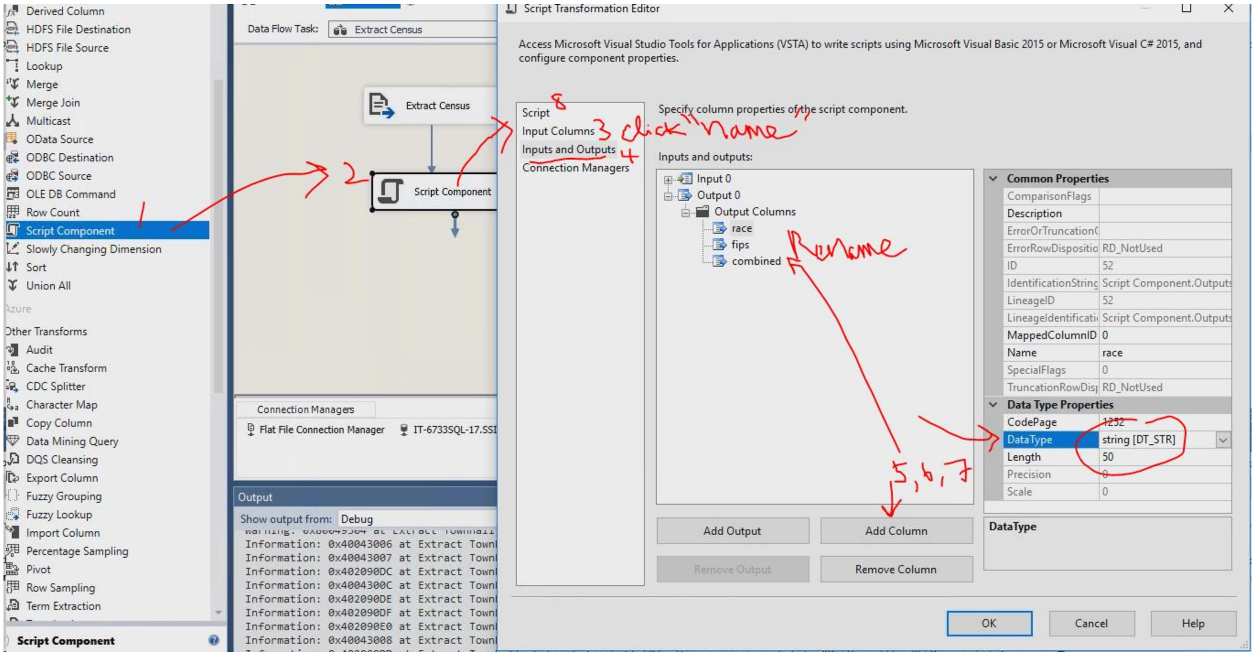


Configure the properties used to connect to and obtain data from a text file.

External Column	Output Column
state_abbr	state_abbr
state_fips	state_fips
county_fips	county_fips
county_name	county_name
fips_class_code	fips_class_code

```
Information: 0x40043006 at Extract Townh
Information: 0x40043007 at Extract Townh
Information: 0x402090DC at Extract Townh
Information: 0x4004300C at Extract Townh
Information: 0x402090DE at Extract Townh
Information: 0x402090DF at Extract Townh
Information: 0x402090E0 at Extract Townh
Information: 0x40043008 at Extract Townh
Information: 0x402090DD at Extract Townhall, Extract Townhall [2]: The processing of file "D:\Users\tkim25\Documents\abe exar
```

35. Create a Script Component to add “combined” column.



Data Flow Task: Extract Census

Connection Managers: Flat File Connection Manager; IT-6733SQL-17.SSI

Output

Show output from: Debug

```
Information: 0x40043006 at Extract Town
Information: 0x40043007 at Extract Town
Information: 0x402090DC at Extract Town
Information: 0x4004300C at Extract Town
Information: 0x402090DE at Extract Town
Information: 0x402090DF at Extract Town
Information: 0x402090E0 at Extract Town
Information: 0x40043008 at Extract Town
```

Access Microsoft Visual Studio Tools for Applications (VSTA) to write scripts using Microsoft Visual Basic 2015 or Microsoft Visual C# 2015, and configure component properties.

Specify column properties of the script component.

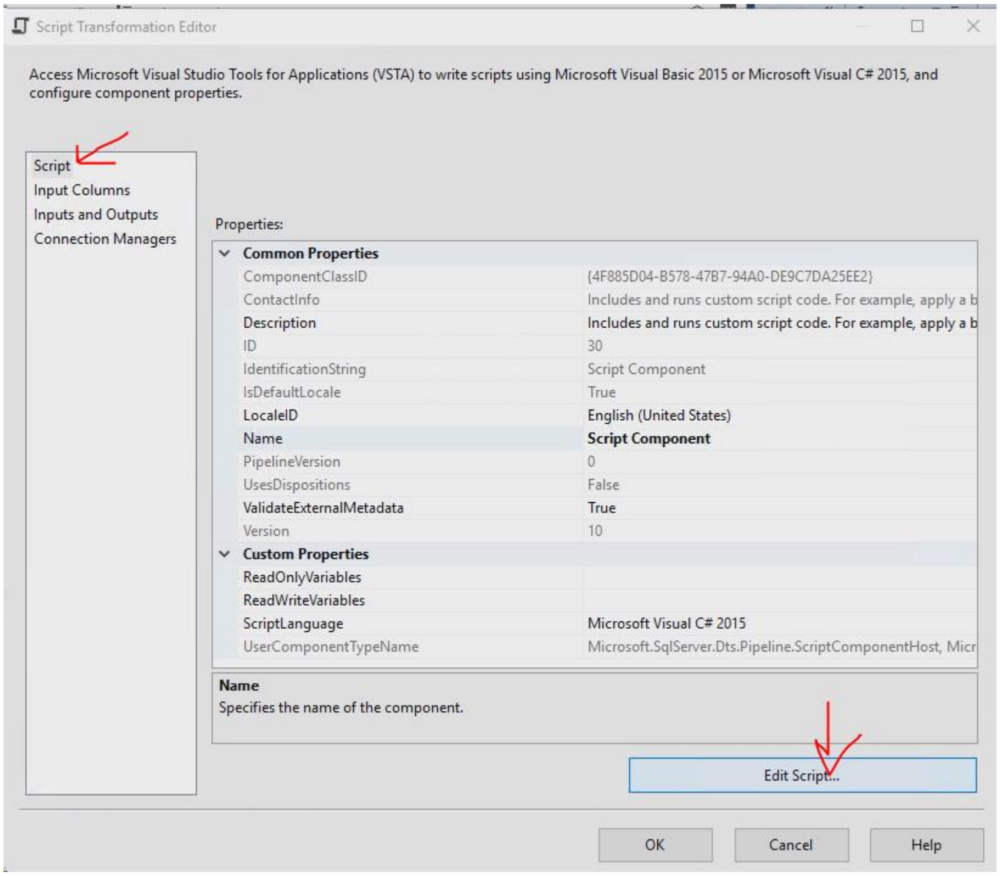
Inputs and outputs:

- Input 0
- Output 0
  - Output Columns
    - race
    - fips
    - combined

Common Properties	
ComparisonFlags	
Description	
ErrorOrTruncation(	
ErrorRowDispositio	RD_NotUsed
ID	52
IdentificationString	Script Component.Outputs
LineageID	52
LineageIdentificati	Script Component.Outputs
MappedColumnID	0
Name	race
SpecialFlags	0
TruncationRowDis	RD_NotUsed
Data Type Properties	
CodePage	1252
DataType	string [DT_STR]
Length	50
Precision	0
Scale	0

DataType

36. Script → Edit Script



Access Microsoft Visual Studio Tools for Applications (VSTA) to write scripts using Microsoft Visual Basic 2015 or Microsoft Visual C# 2015, and configure component properties.

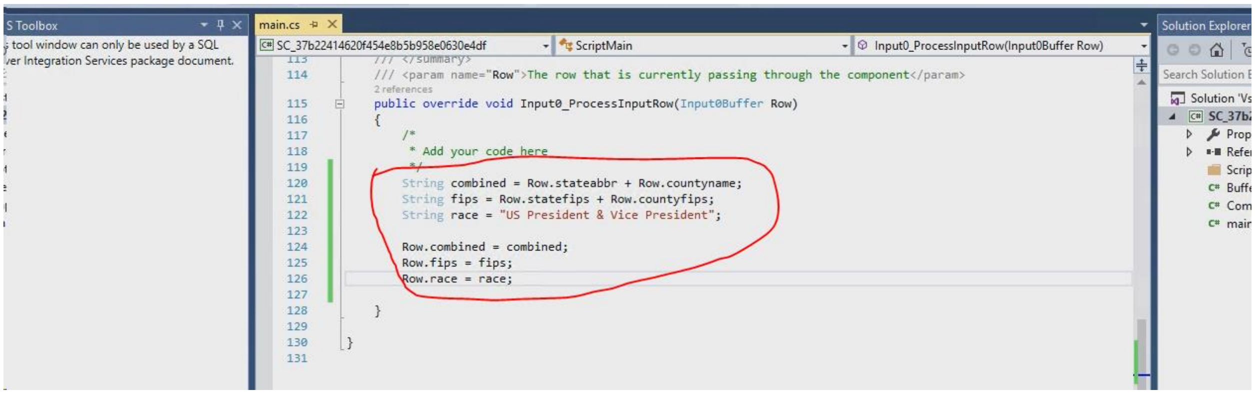
Properties:

Common Properties	
ComponentClassID	{4F885D04-B578-47B7-94A0-DE9C7DA25EE2}
ContactInfo	Includes and runs custom script code. For example, apply a b
Description	Includes and runs custom script code. For example, apply a b
ID	30
IdentificationString	Script Component
IsDefaultLocale	True
LocaleID	English (United States)
Name	**Script Component**
PipelineVersion	0
UsesDispositions	False
ValidateExternalMetadata	True
Version	10
Custom Properties	
ReadOnlyVariables	
ReadWriteVariables	
ScriptLanguage	Microsoft Visual C# 2015
UserComponentTypeName	Microsoft.SqlServer.Dts.Pipeline.ScriptComponentHost, Micr

Name
Specifies the name of the component.

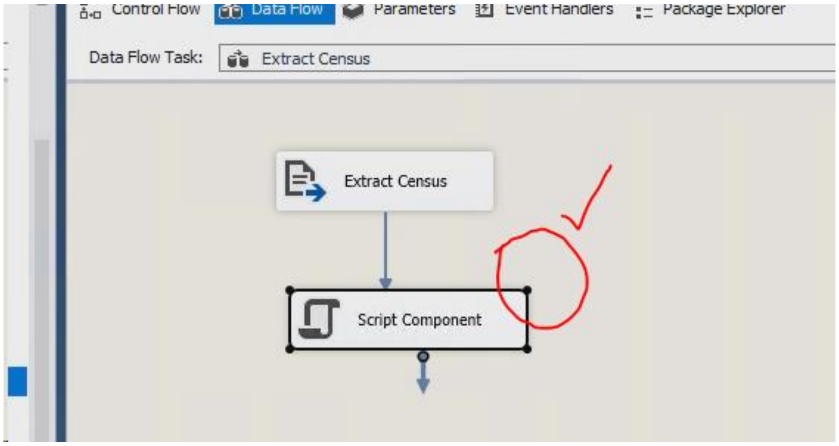
37. Add code at the end for three additional columns (combined, fips, race):

```
String combined = Row.stateabbr + Row.countyname;
String fips = Row.statefips = Row.countyfips;
String race = “US Presidential & Vice President”;

Row.combined = combined;
Row.fips = fips;
Row.race = race;
```


S Toolbox
tool window can only be used by a SQL
ver Integration Services package document.
main.cs
SC_37b22414620f454e8b5b958e0630e4df
ScriptMain
Input0_ProcessInputRow(Input0Buffer Row)
Solution Explorer
Search Solution E
Solution 'Vs
SC_37b
Prop
Refe
Scrip
Buffe
Com
mair

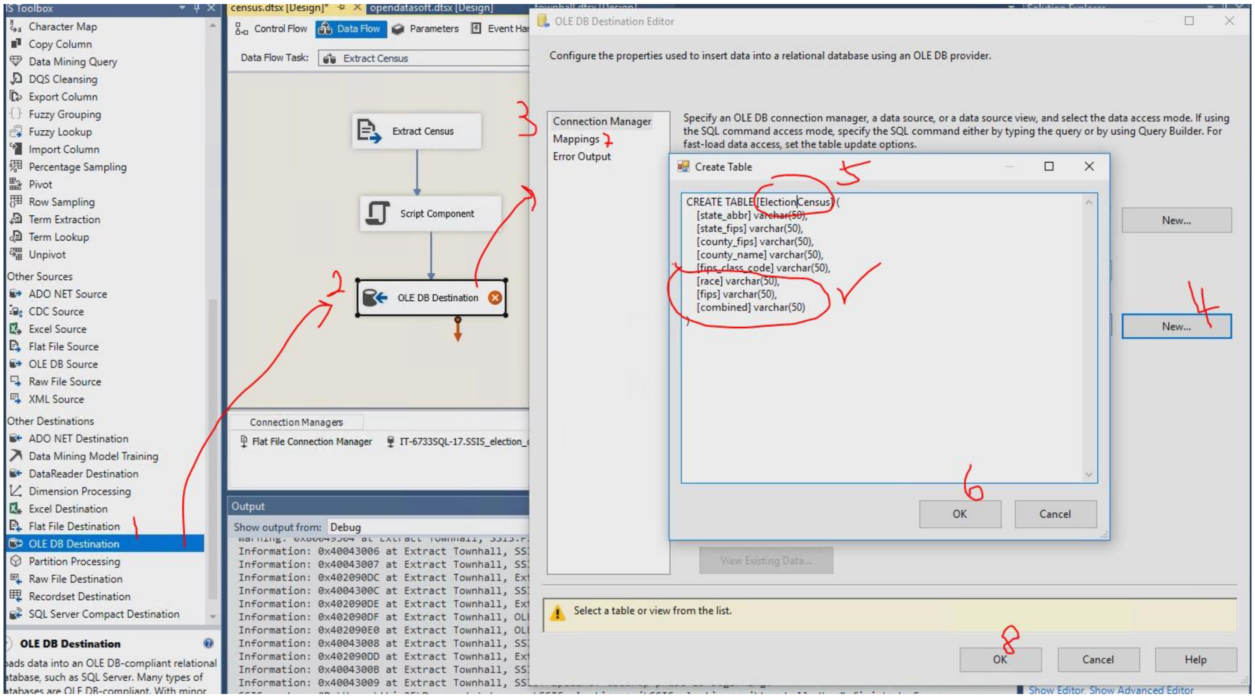
```
        /// </summary>
        /// <param name="Row">The row that is currently passing through the component</param>
        2 references
        public override void Input0_ProcessInputRow(Input0Buffer Row)
        {
            /*
             * Add your code here
             */
            String combined = Row.stateabbr + Row.countyname;
            String fips = Row.statefips + Row.countyfips;
            String race = "US President & Vice President";

            Row.combined = combined;
            Row.fips = fips;
            Row.race = race;

        }

    }
```

[NOTE: Column “race” was manually added into the Census Destination table from Data.gov source.]



Data Flow Task: Extract Census

38. Create OLE DB Destination, rename destination table “ElectionCensus”, check mappings.



OLE DB Destination

…oads data into an OLE DB-compliant relational …atabase, such as SQL Server. Many types of …atabases are OLE DB-compliant. With minor

Configure the properties used to insert data into a relational database using an OLE DB provider.

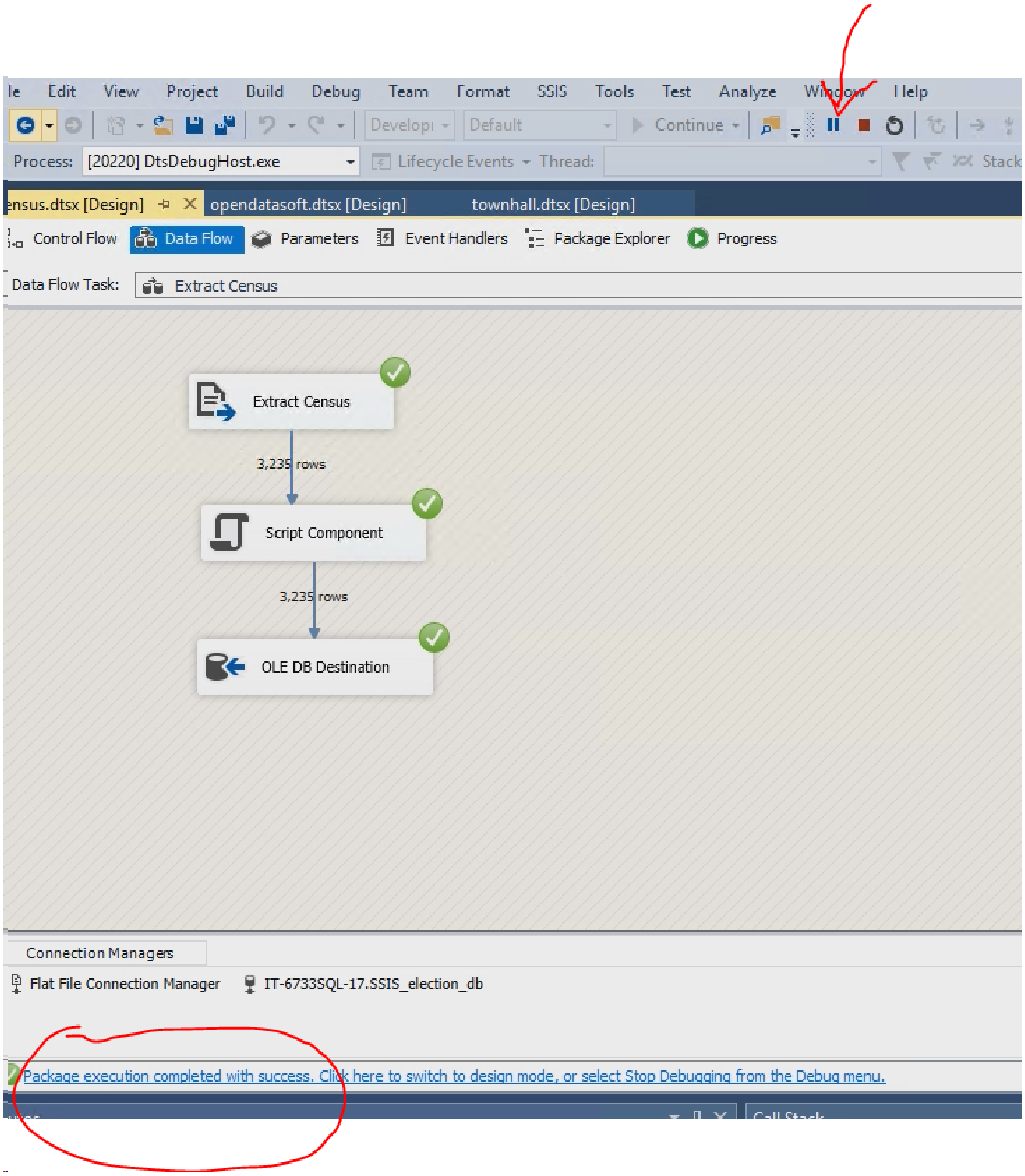
Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

```
CREATE TABLE [ElectionCensus] (
    [state_abbr] varchar(50),
    [state_fips] varchar(50),
    [county_fips] varchar(50),
    [county_name] varchar(50),
    [fips_class_code] varchar(50),
    [race] varchar(50),
    [fips] varchar(50),
    [combined] varchar(50)
)
```

⚠ Select a table or view from the list.

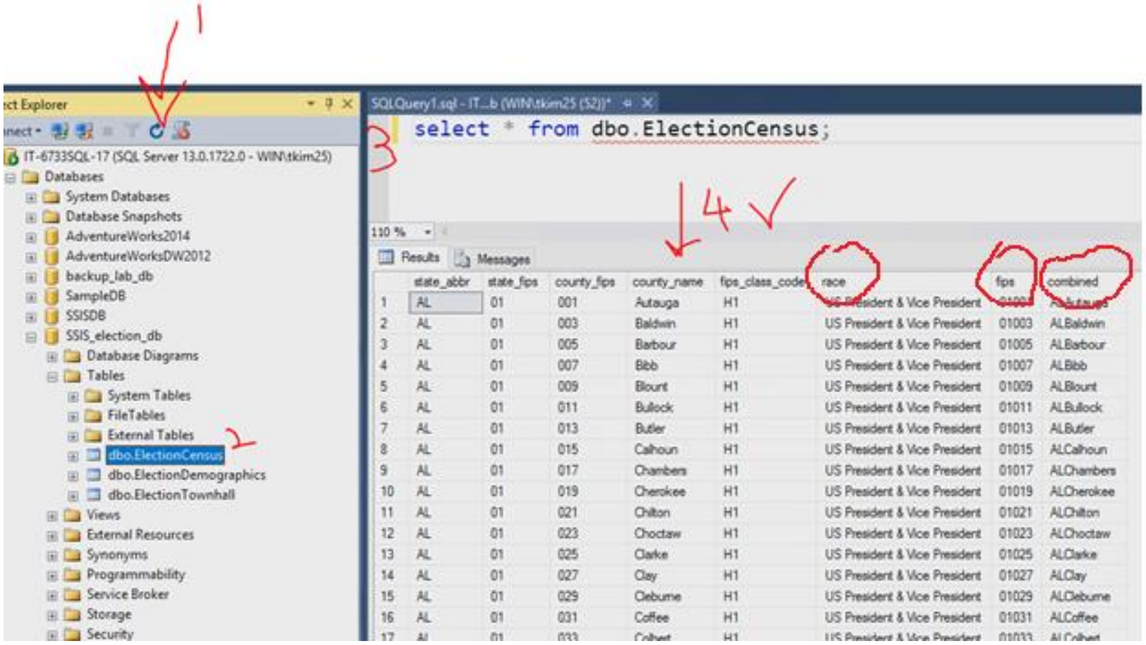
```
Information: 0x40043006 at Extract Townhall, SS
Information: 0x40043007 at Extract Townhall, SS
Information: 0x402090DC at Extract Townhall, Ext
Information: 0x4004300C at Extract Townhall, SS
Information: 0x402090DE at Extract Townhall, Ext
Information: 0x402090DF at Extract Townhall, OL
Information: 0x402090E0 at Extract Townhall, OL
Information: 0x40043008 at Extract Townhall, SS
Information: 0x402090DD at Extract Townhall, Ext
Information: 0x4004300B at Extract Townhall, SS
Information: 0x40043009 at Extract Townhall, SS
```

39. Click start, check successful execution.



Package execution completed with success. Click here to switch to design mode, or select Stop Debugging from the Debug menu.

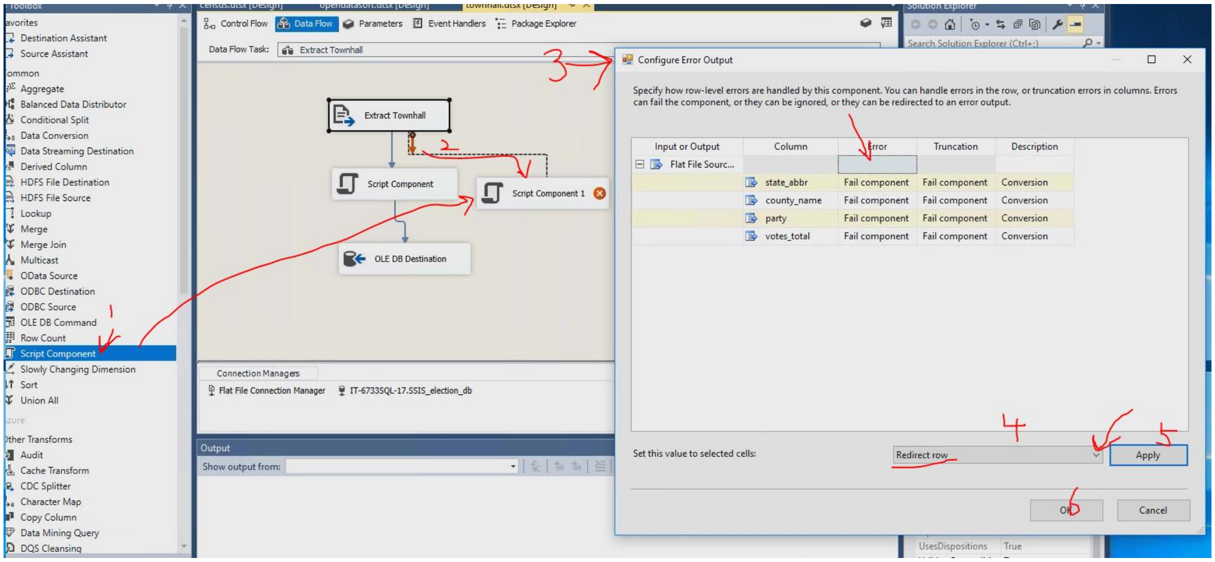
40. Go to SQL Server Management Studio, check that new destination table is added, run test query, check that newly added columns are there.



```
select * from dbo.ElectionCensus;
```

	state_abbr	state_fips	county_fips	county_name	fips_class_code	race	fips	combined
1	AL	01	001	Autauga	H1	US President & Vice President	01001	ALAutauga
2	AL	01	003	Baldwin	H1	US President & Vice President	01003	ALBaldwin
3	AL	01	005	Barbour	H1	US President & Vice President	01005	ALBarbour
4	AL	01	007	Bibb	H1	US President & Vice President	01007	ALBibb
5	AL	01	009	Blount	H1	US President & Vice President	01009	ALBlount
6	AL	01	011	Bullock	H1	US President & Vice President	01011	ALBullock
7	AL	01	013	Butler	H1	US President & Vice President	01013	ALButler
8	AL	01	015	Calhoun	H1	US President & Vice President	01015	ALCalhoun
9	AL	01	017	Chambers	H1	US President & Vice President	01017	ALChambers
10	AL	01	019	Cherokee	H1	US President & Vice President	01019	ALCherokee
11	AL	01	021	Chilton	H1	US President & Vice President	01021	ALChilton
12	AL	01	023	Choctaw	H1	US President & Vice President	01023	ALChoctaw
13	AL	01	025	Clarke	H1	US President & Vice President	01025	ALClarke
14	AL	01	027	Clay	H1	US President & Vice President	01027	ALClay
15	AL	01	029	Cleburne	H1	US President & Vice President	01029	ALCleburne
16	AL	01	031	Coffee	H1	US President & Vice President	01031	ALCoffee
17	AL	01	033	Colbert	H1	US President & Vice President	01033	ALColbert

41. Create Script Component for Error Description.



Configure Error Output

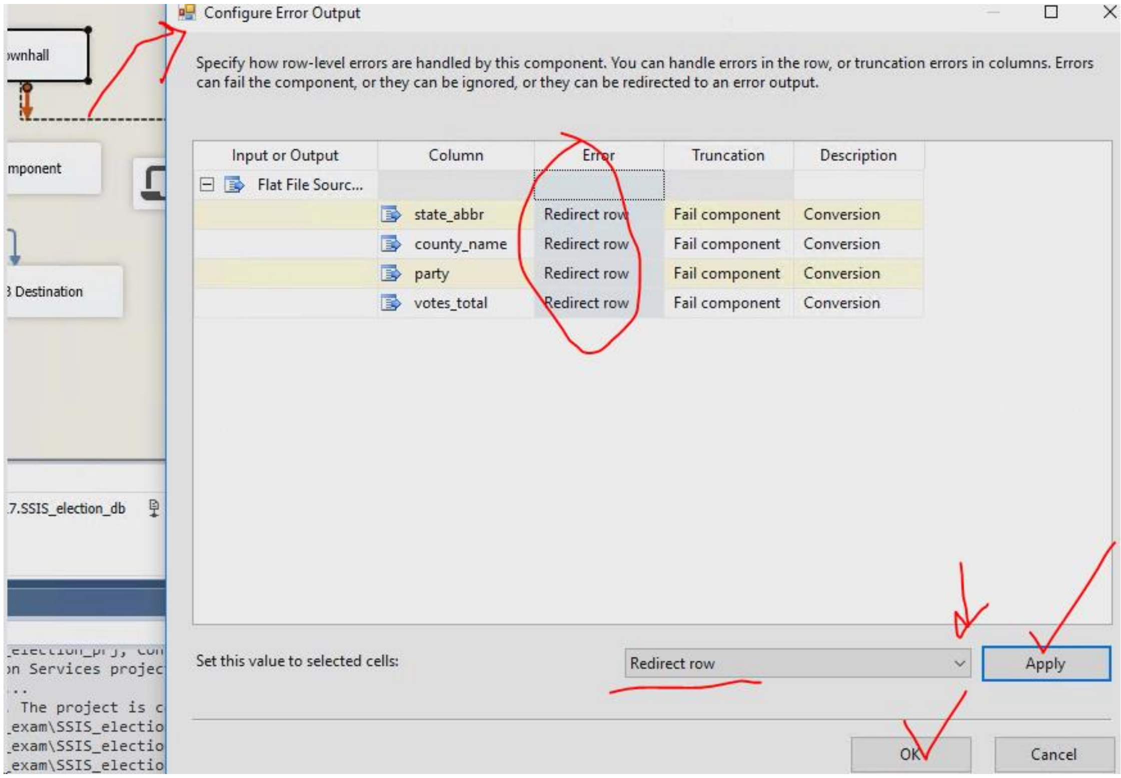
Specify how row-level errors are handled by this component. You can handle errors in the row, or truncation errors in columns. Errors can fail the component, or they can be ignored, or they can be redirected to an error output.

Input or Output	Column	Error	Truncation	Description
⊟ Flat File Sourc...				
	state_abbr	Fail component	Fail component	Conversion
	county_name	Fail component	Fail component	Conversion
	party	Fail component	Fail component	Conversion
	votes_total	Fail component	Fail component	Conversion

Set this value to selected cells: Redirect row

Apply

OK Cancel

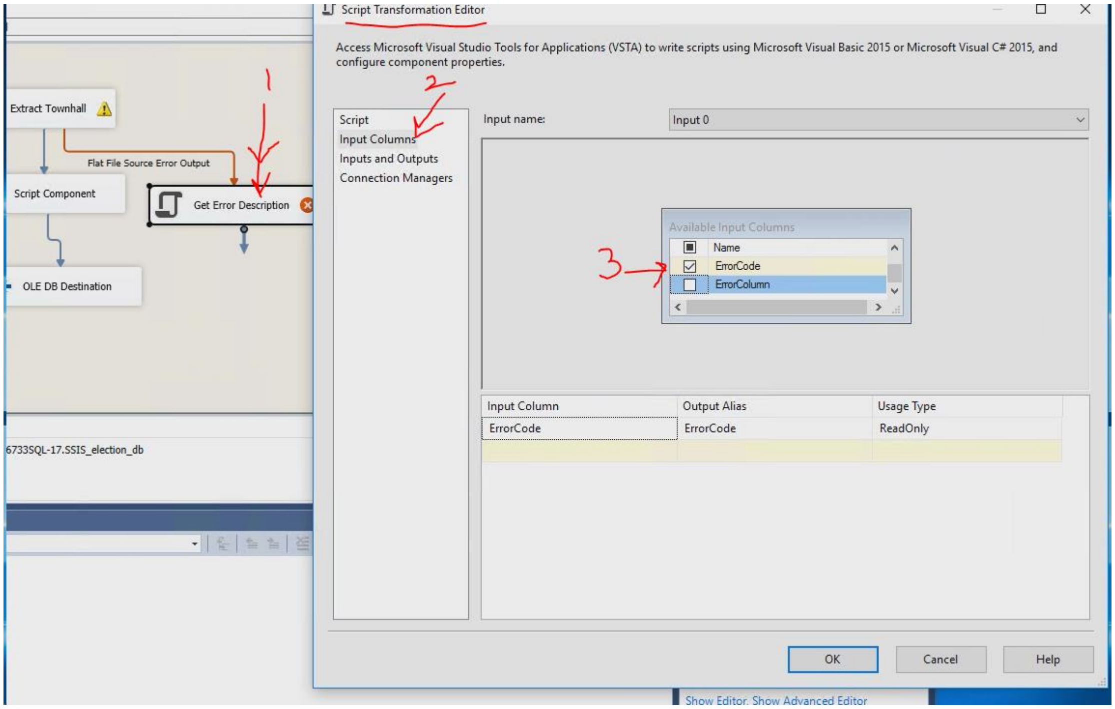


Configure Error Output

Specify how row-level errors are handled by this component. You can handle errors in the row, or truncation errors in columns. Errors can fail the component, or they can be ignored, or they can be redirected to an error output.

Input or Output	Column	Error	Truncation	Description
⊟ Flat File Sourc...				
	state_abbr	Redirect row	Fail component	Conversion
	county_name	Redirect row	Fail component	Conversion
	party	Redirect row	Fail component	Conversion
	votes_total	Redirect row	Fail component	Conversion

Set this value to selected cells: Redirect row



Script Transformation Editor

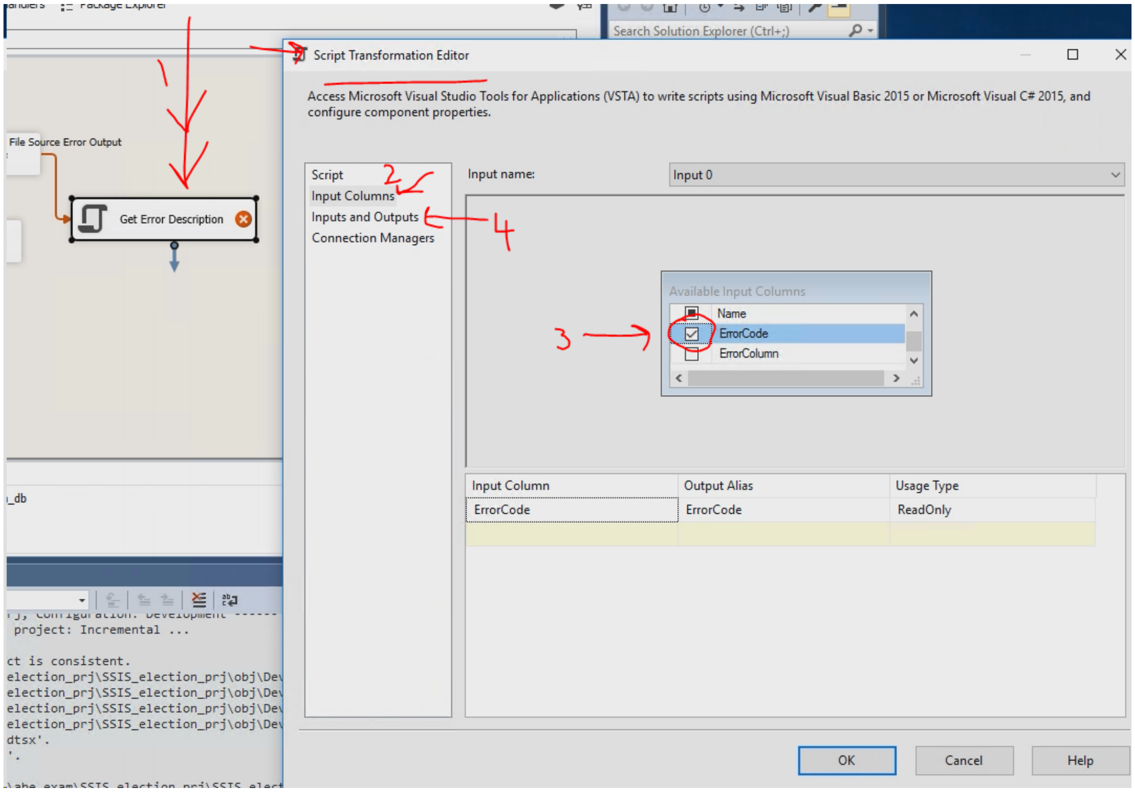
Access Microsoft Visual Studio Tools for Applications (VSTA) to write scripts using Microsoft Visual Basic 2015 or Microsoft Visual C# 2015, and configure component properties.

Input name: Input 0

Available Input Columns

■	Name
☑	ErrorCode
☐	ErrorColumn

Input Column	Output Alias	Usage Type
ErrorCode	ErrorCode	ReadOnly



Script Transformation Editor

Access Microsoft Visual Studio Tools for Applications (VSTA) to write scripts using Microsoft Visual Basic 2015 or Microsoft Visual C# 2015, and configure component properties.

- Script
- Input Columns
- Inputs and Outputs
- Connection Managers

Input name: Input 0

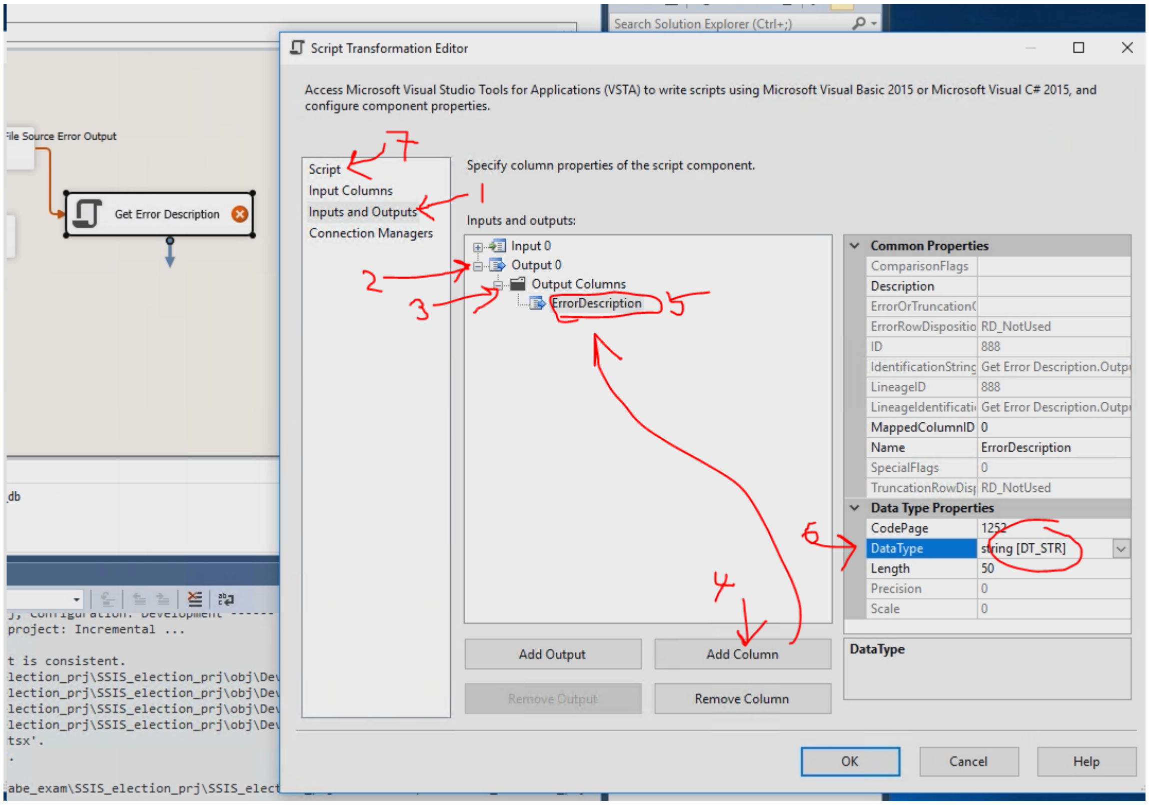
Available Input Columns

- ☐ Name
- ☑ ErrorCode
- ☐ ErrorColumn

Input Column	Output Alias	Usage Type
ErrorCode	ErrorCode	ReadOnly

OK | Cancel | Help

```
project: Incremental ...

ct is consistent.
election_prj\SSIS_election_prj\obj\Dev
election_prj\SSIS_election_prj\obj\Dev
election_prj\SSIS_election_prj\obj\Dev
election_prj\SSIS_election_prj\obj\Dev
dtsx'.
'.
```



Script Transformation Editor

Access Microsoft Visual Studio Tools for Applications (VSTA) to write scripts using Microsoft Visual Basic 2015 or Microsoft Visual C# 2015, and configure component properties.

- Script
- Input Columns
- Inputs and Outputs
- Connection Managers

Specify column properties of the script component.

Inputs and outputs:

- Input 0
- Output 0
  - Output Columns
    - ErrorDescription

Common Properties	
ComparisonFlags	
Description	
ErrorOrTruncation(	
ErrorRowDispositio	RD_NotUsed
ID	888
IdentificationString	Get Error Description.Outp
LineageID	888
LineageIdentificati	Get Error Description.Outp
MappedColumnID	0
Name	ErrorDescription
SpecialFlags	0
TruncationRowDis	RD_NotUsed
Data Type Properties	
CodePage	1252
DataType	string [DT_STR]
Length	50
Precision	0
Scale	0

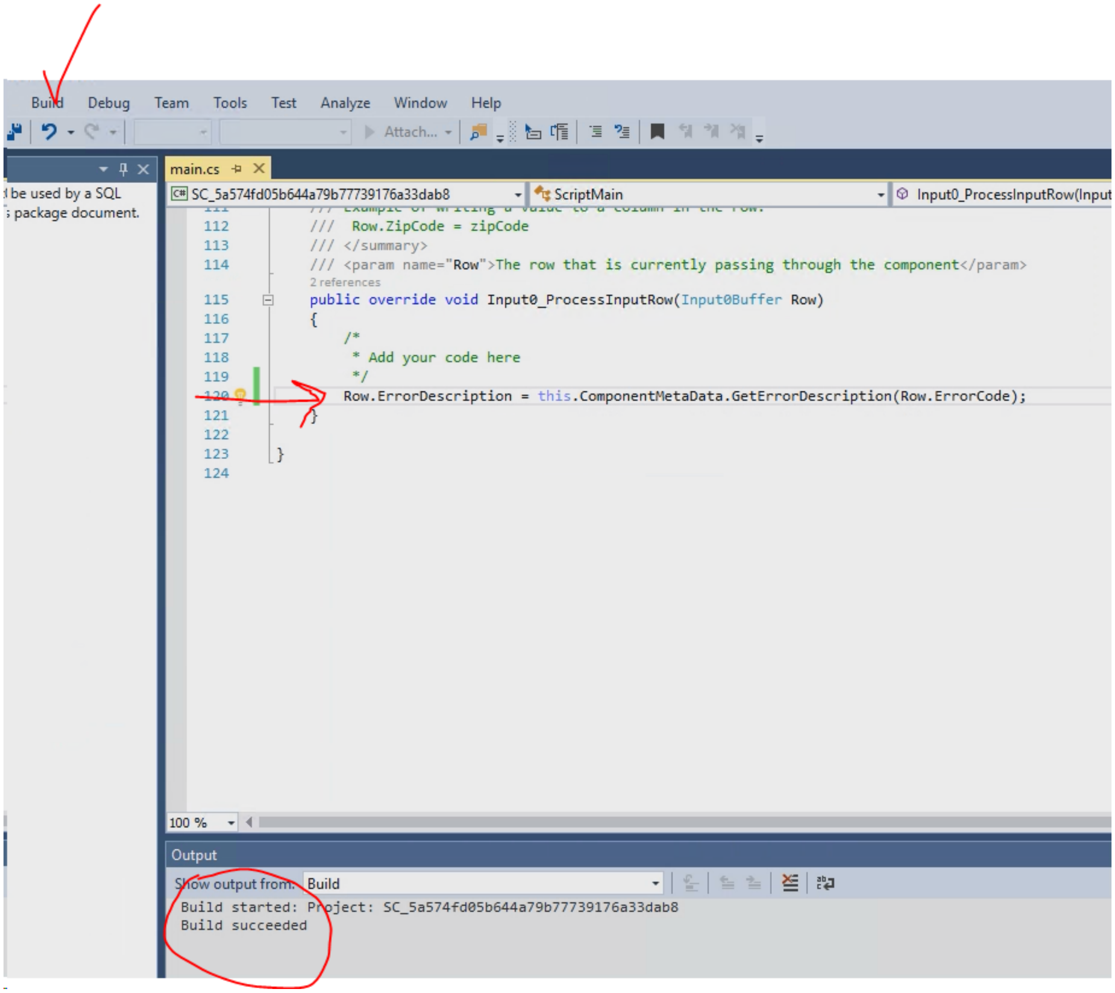
DataType

Add Output | Add Column | Remove Output | Remove Column

OK | Cancel | Help

```
project: Incremental ...

t is consistent.
lection_prj\SSIS_election_prj\obj\Dev
lection_prj\SSIS_election_prj\obj\Dev
lection_prj\SSIS_election_prj\obj\Dev
lection_prj\SSIS_election_prj\obj\Dev
tsx'.
.
abe_exam\SSIS_election_prj\SSIS_elect
```

In Script, click Edit Script, then add code, then click Build as shown on next screenshot:



be used by a SQL
package document.

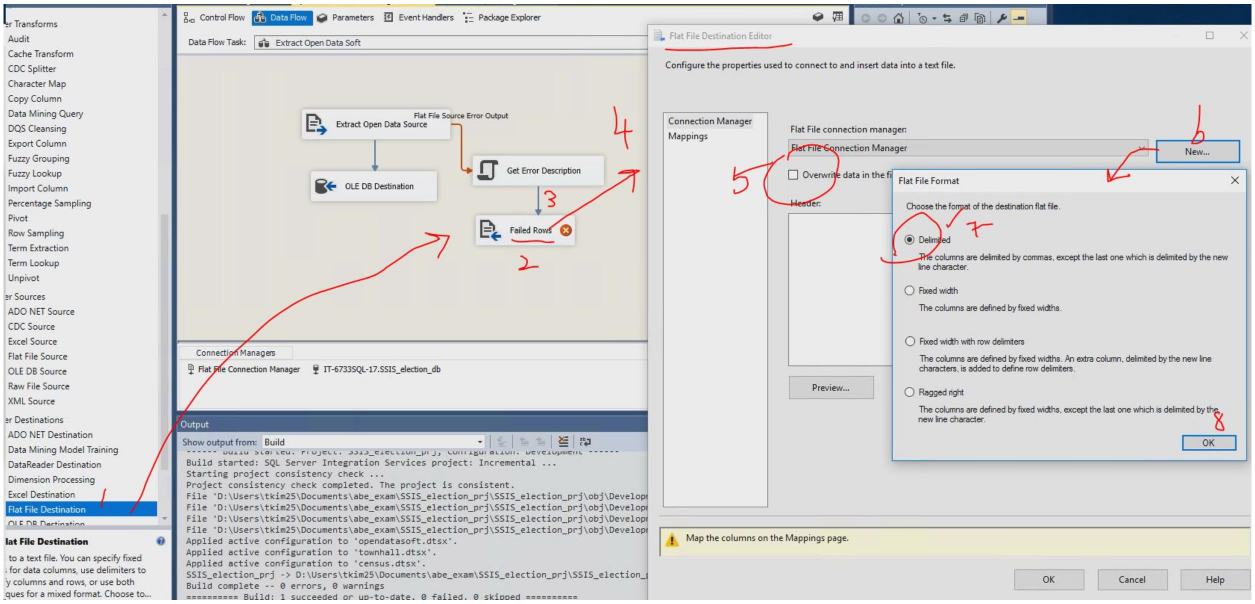
```
    /// Example of writing a value to a column in the row:
    ///   Row.ZipCode = zipCode
    /// </summary>
    /// <param name="Row">The row that is currently passing through the component</param>
    2 references
    public override void Input0_ProcessInputRow(Input0Buffer Row)
    {
        /*
         * Add your code here
         */
        Row.ErrorDescription = this.ComponentMetaData.GetErrorDescription(Row.ErrorCode);
    }

}
```

Output

Show output from: Build

```
Build started: Project: SC_5a574fd05b644a79b77739176a33dab8
Build succeeded
```

Then create Flat File Destination to receive error description:



lat File Destination

to a text file. You can specify fixed
for data columns, use delimiters to
y columns and rows, or use both
ques for a mixed format. Choose to...

Output

Show output from: Build

```
------ Build started: Project: SSIS_election_prj, Configuration: Development ------
Build started: SQL Server Integration Services project: Incremental ...
Starting project consistency check ...
Project consistency check completed. The project is consistent.
File 'D:\Users\tkim25\Documents\abe_exam\SSIS_election_prj\SSIS_election_prj\obj\Developm
File 'D:\Users\tkim25\Documents\abe_exam\SSIS_election_prj\SSIS_election_prj\obj\Developm
File 'D:\Users\tkim25\Documents\abe_exam\SSIS_election_prj\SSIS_election_prj\obj\Developm
File 'D:\Users\tkim25\Documents\abe_exam\SSIS_election_prj\SSIS_election_prj\obj\Developm
Applied active configuration to 'opendatasoft.dtsx'.
Applied active configuration to 'townhall.dtsx'.
Applied active configuration to 'census.dtsx'.
SSIS_election_prj -> D:\Users\tkim25\Documents\abe_exam\SSIS_election_prj\SSIS_election_
Build complete -- 0 errors, 0 warnings
========== Build: 1 succeeded or up-to-date, 0 failed, 0 skipped ==========
```

Configure the properties used to connect to and insert data into a text file.

Flat File connection manager: Flat File Connection Manager

□ Overwrite data in the fi

Header:

Flat File Format

Choose the format of the destination flat file.

● Delimited

The columns are delimited by commas, except the last one which is delimited by the new line character.

○ Fixed width

The columns are defined by fixed widths.

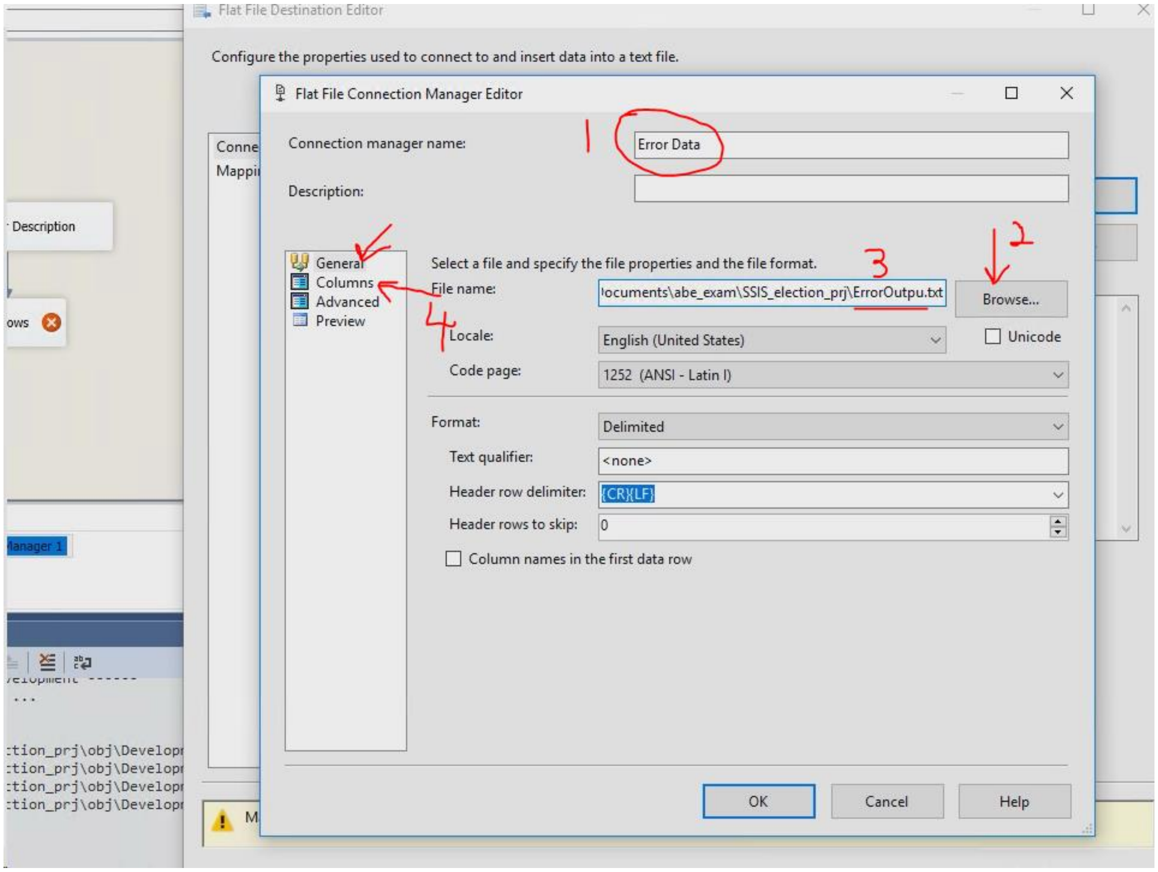
○ Fixed width with row delimiters

The columns are defined by fixed widths. An extra column, delimited by the new line characters, is added to define row delimiters.

○ Ragged right

The columns are defined by fixed widths, except the last one which is delimited by the new line character.

Map the columns on the Mappings page.



Configure the properties used to connect to and insert data into a text file.

Flat File Connection Manager Editor

Connection manager name: Error Data

Description:

Select a file and specify the file properties and the file format.

File name: !ocuments\abe_exam\SSIS_election_prj\ErrorOutpu.txt

Locale: English (United States)

□ Unicode

Code page: 1252 (ANSI - Latin I)

Format: Delimited

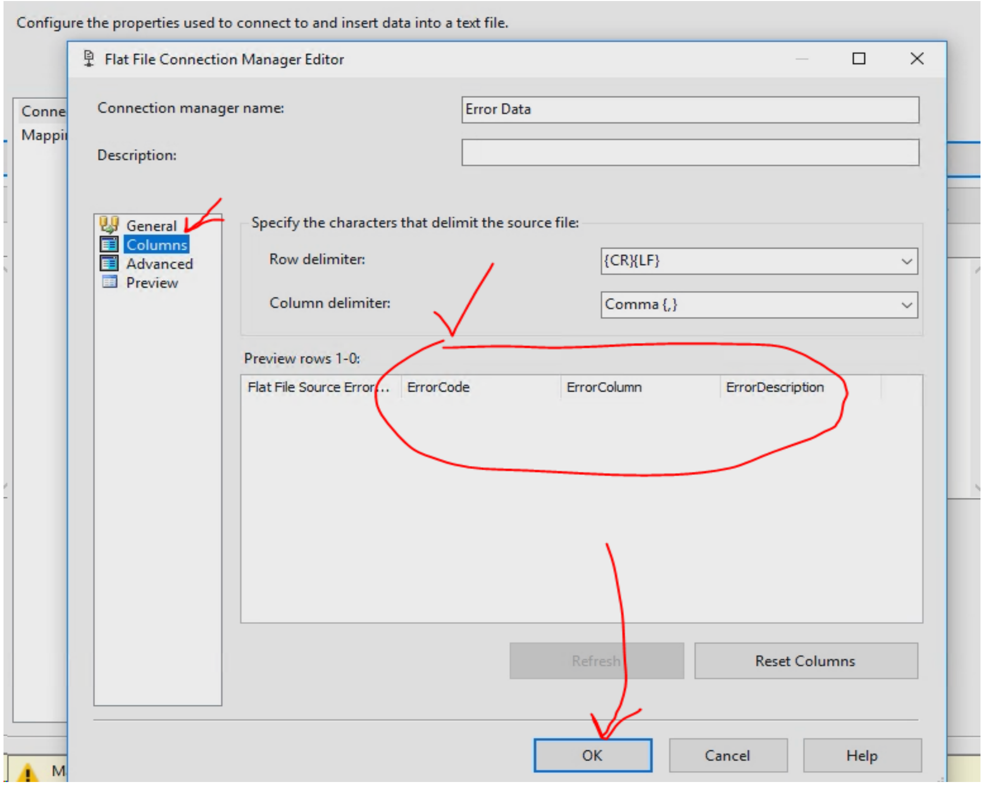
Text qualifier: <none>

Header row delimiter: {CR}{LF}

Header rows to skip: 0

□ Column names in the first data row

```
:tion_prj\obj\Develop
:tion_prj\obj\Develop
:tion_prj\obj\Develop
:tion_prj\obj\Develop
```



Configure the properties used to connect to and insert data into a text file.

Flat File Connection Manager Editor

Connection manager name: Error Data

Description:

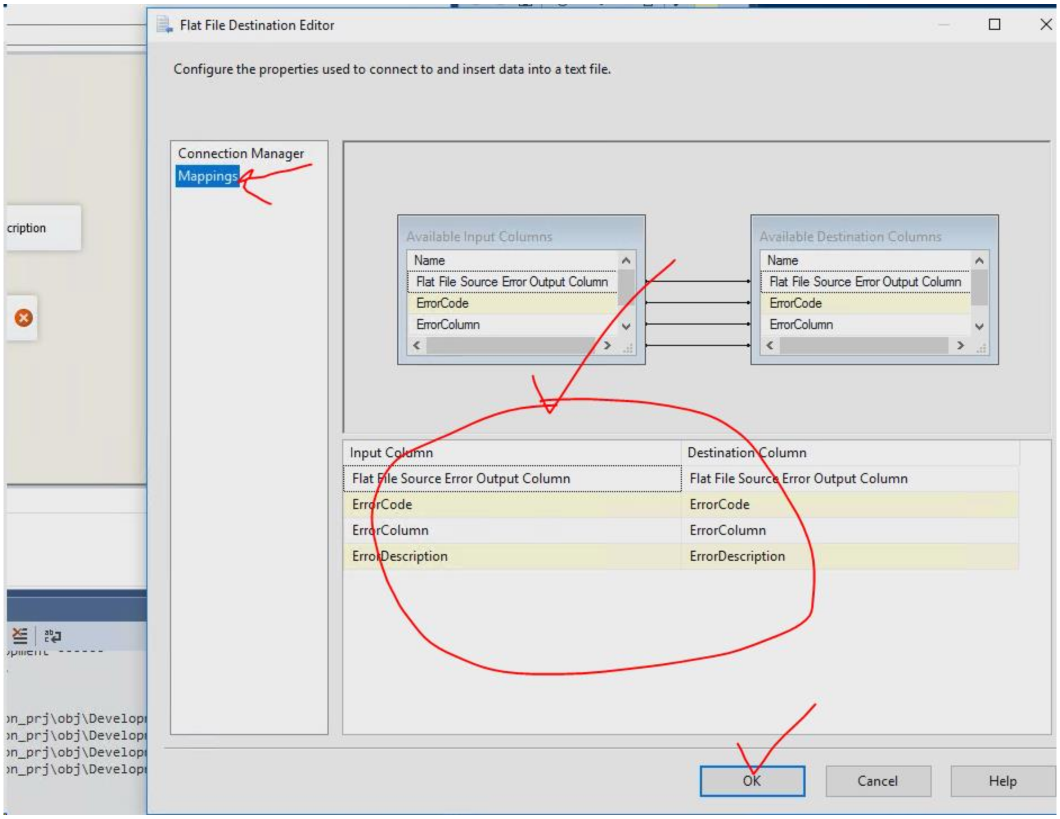
Specify the characters that delimit the source file:

Row delimiter: {CR}{LF}

Column delimiter: Comma {,}

Preview rows 1-0:

Flat File Source Error...	ErrorCode	ErrorColumn	ErrorDescription

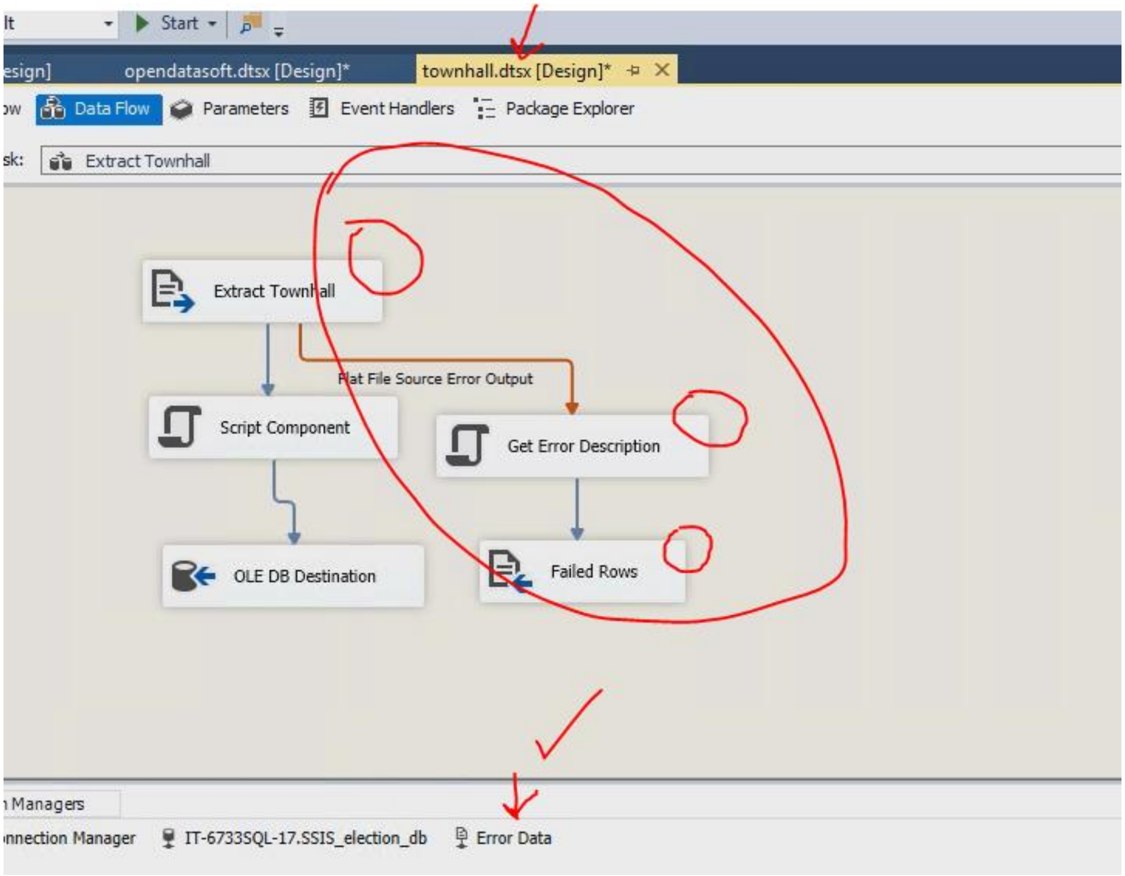


Flat File Destination Editor

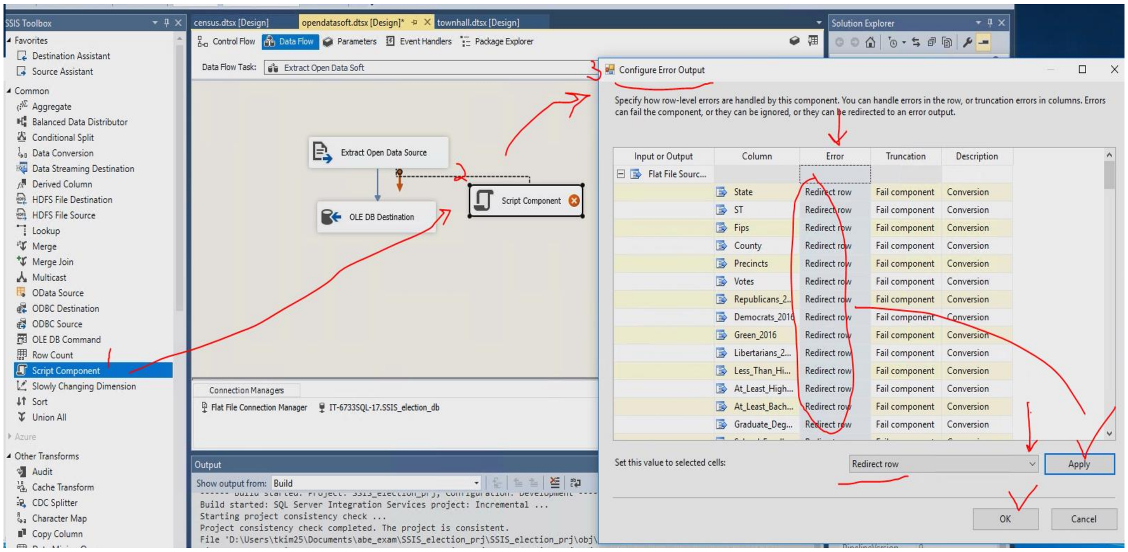
Configure the properties used to connect to and insert data into a text file.

Input Column	Destination Column
Flat File Source Error Output Column	Flat File Source Error Output Column
ErrorCode	ErrorCode
ErrorColumn	ErrorColumn
ErrorDescription	ErrorDescription

```
ployment ------
on_prj\obj\Develop
on_prj\obj\Develop
on_prj\obj\Develop
on_prj\obj\Develop
```



42. Repeat steps, create Get Error Description for OpenDataSoft:



Data Flow Task: Extract Open Data Soft

```
------ Build started: Project: SSIS_election_prj, Configuration: Development ------
Build started: SQL Server Integration Services project: Incremental ...
Starting project consistency check ...
Project consistency check completed. The project is consistent.
File 'D:\Users\tkim25\Documents\abe_exam\SSIS_election_prj\SSIS_election_prj\obj\
```

Configure Error Output

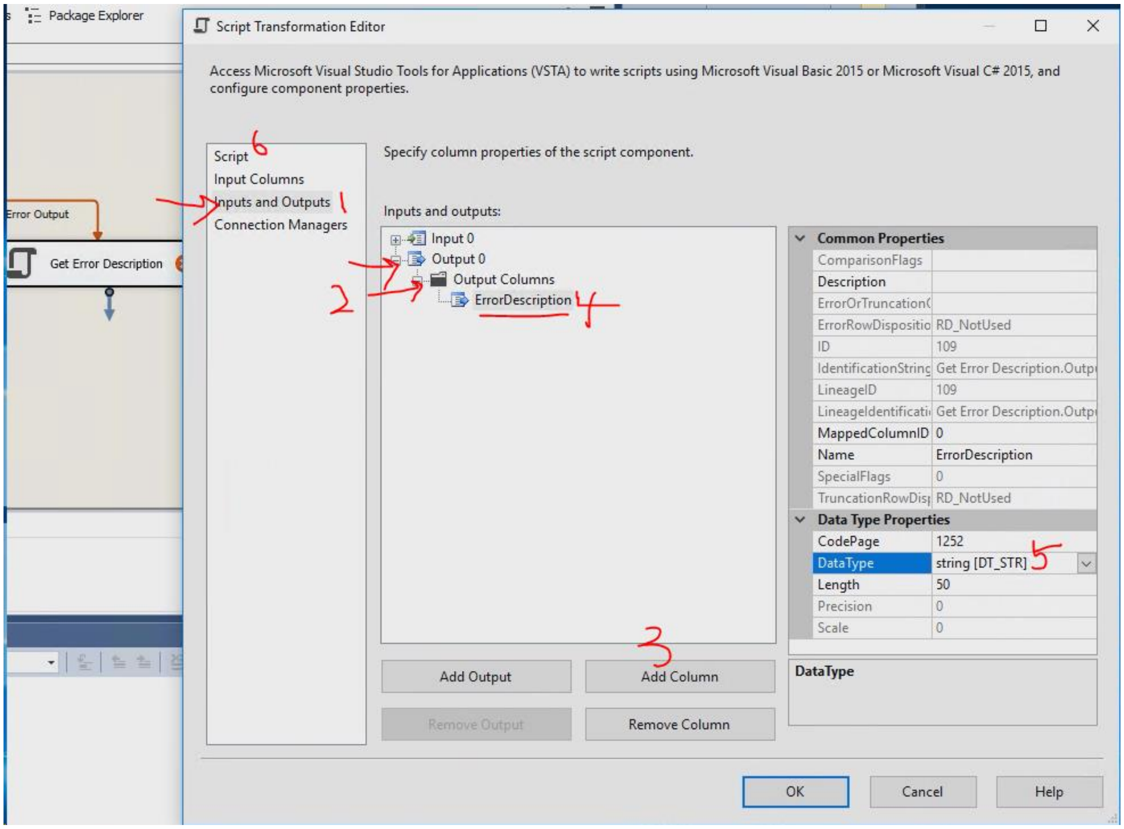
Specify how row-level errors are handled by this component. You can handle errors in the row, or truncation errors in columns. Errors can fail the component, or they can be redirected to an error output.

Input or Output	Column	Error	Truncation	Description
Flat File Sourc...				
	State	Redirect row	Fail component	Conversion
	ST	Redirect row	Fail component	Conversion
	Fips	Redirect row	Fail component	Conversion
	County	Redirect row	Fail component	Conversion
	Precincts	Redirect row	Fail component	Conversion
	Votes	Redirect row	Fail component	Conversion
	Republicans_2...	Redirect row	Fail component	Conversion
	Democrats_2016	Redirect row	Fail component	Conversion
	Green_2016	Redirect row	Fail component	Conversion
	Libertarians_2...	Redirect row	Fail component	Conversion
	Less_Than_Hi...	Redirect row	Fail component	Conversion
	At_Least_High...	Redirect row	Fail component	Conversion
	At_Least_Bach...	Redirect row	Fail component	Conversion
	Graduate_Deg...	Redirect row	Fail component	Conversion

Set this value to selected cells: Redirect row

Apply

OK Cancel



Script Transformation Editor

Access Microsoft Visual Studio Tools for Applications (VSTA) to write scripts using Microsoft Visual Basic 2015 or Microsoft Visual C# 2015, and configure component properties.

- Script
- Input Columns
- Inputs and Outputs
- Connection Managers

Specify column properties of the script component.

Inputs and outputs:

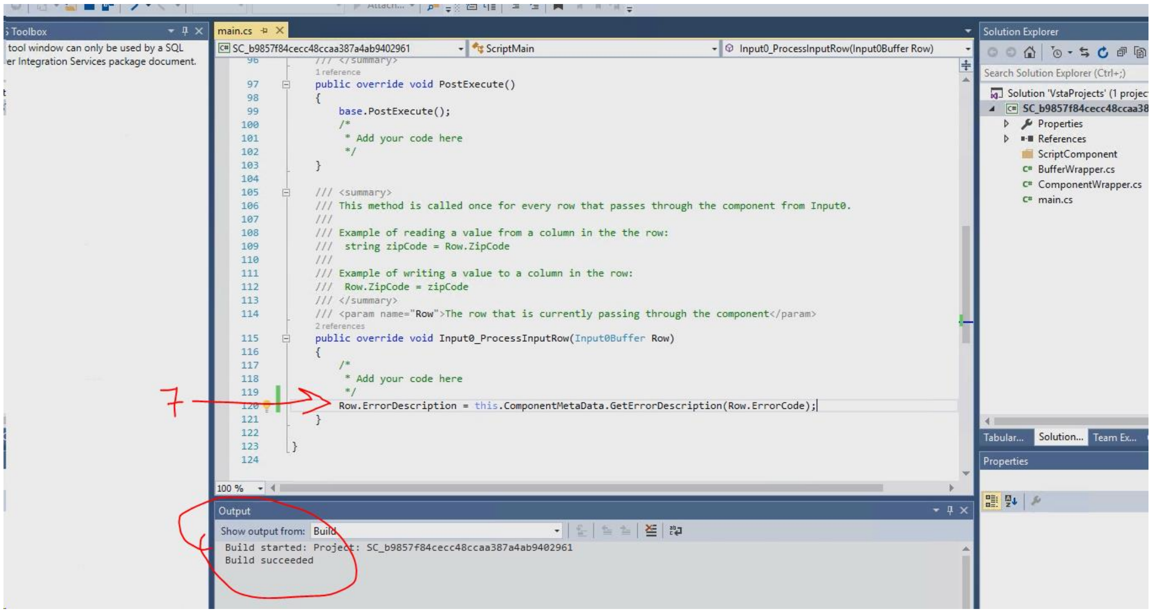
- Input 0
- Output 0
  - Output Columns
    - ErrorDescription

Common Properties	
ComparisonFlags	
Description	
ErrorOrTruncation(	
ErrorRowDispositio	RD_NotUsed
ID	109
IdentificationString	Get Error Description.Outp
LineageID	109
LineageIdentificati	Get Error Description.Outp
MappedColumnID	0
Name	ErrorDescription
SpecialFlags	0
TruncationRowDis	RD_NotUsed
Data Type Properties	
CodePage	1252
DataType	string [DT_STR]
Length	50
Precision	0
Scale	0

DataType

Add Output | Add Column | Remove Output | Remove Column

OK | Cancel | Help

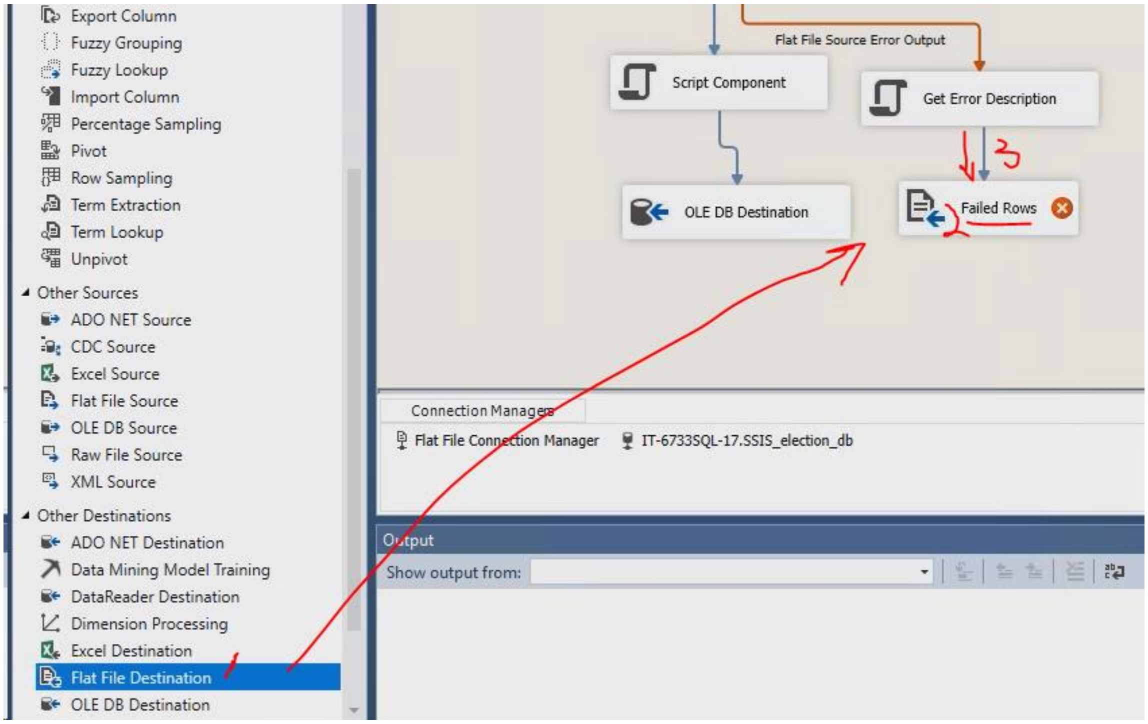

Toolbox
tool window can only be used by a SQL
er Integration Services package document.
main.cs
SC_b9857f84cecc48ccaa387a4ab9402961
ScriptMain
Input0_ProcessInputRow(Input0Buffer Row)
Solution Explorer
Search Solution Explorer (Ctrl+;)
Solution 'VstaProjects' (1 projec
SC_b9857f84cecc48ccaa38
Properties
References
ScriptComponent
BufferWrapper.cs
ComponentWrapper.cs
main.cs
Tabular...
Solution...
Team Ex...
Properties
100 %
7

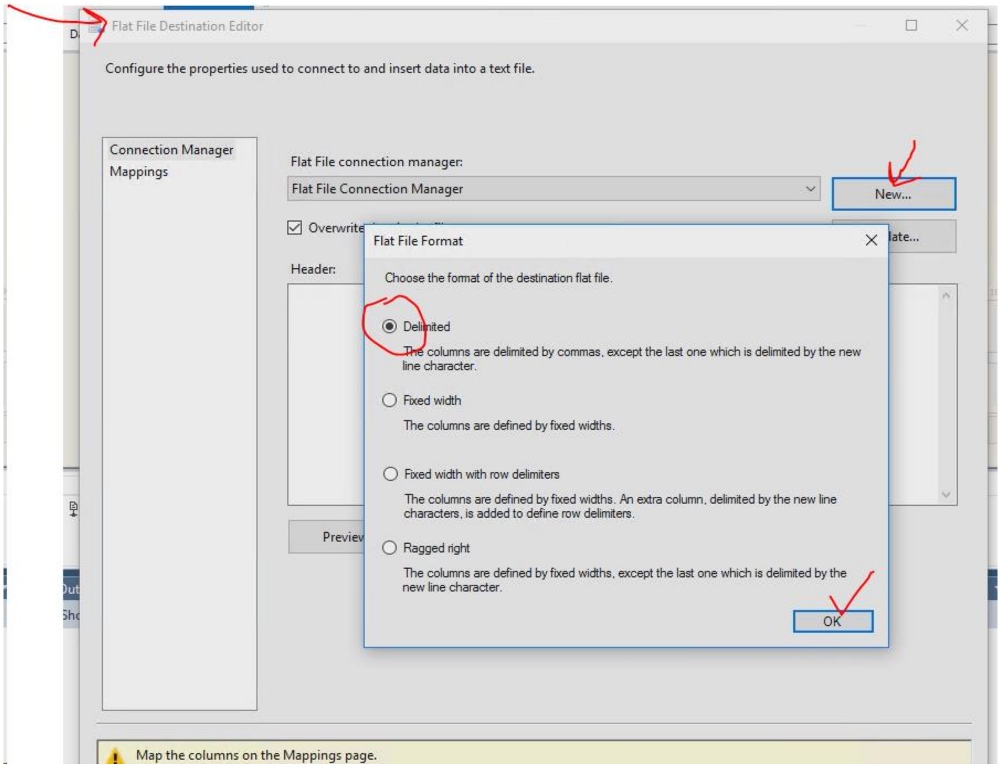
```
        /// </summary>
        public override void PostExecute()
        {
            base.PostExecute();
            /*
             * Add your code here
             */
        }

        /// <summary>
        /// This method is called once for every row that passes through the component from Input0.
        ///
        /// Example of reading a value from a column in the the row:
        ///  string zipCode = Row.ZipCode
        ///
        /// Example of writing a value to a column in the row:
        ///  Row.ZipCode = zipCode
        /// </summary>
        /// <param name="Row">The row that is currently passing through the component</param>
        public override void Input0_ProcessInputRow(Input0Buffer Row)
        {
            /*
             * Add your code here
             */
            Row.ErrorDescription = this.ComponentMetaData.GetErrorDescription(Row.ErrorCode);
        }

    }
```

Output

Show output from: Build

```
Build started: Project: SC_b9857f84cecc48ccaa387a4ab9402961
Build succeeded
```





Flat File Destination Editor

Configure the properties used to connect to and insert data into a text file.

Flat File connection manager: Flat File Connection Manager

☑ Overwrite

Header:

Flat File Format

Choose the format of the destination flat file.

● Delimited

The columns are delimited by commas, except the last one which is delimited by the new line character.

○ Fixed width

The columns are defined by fixed widths.

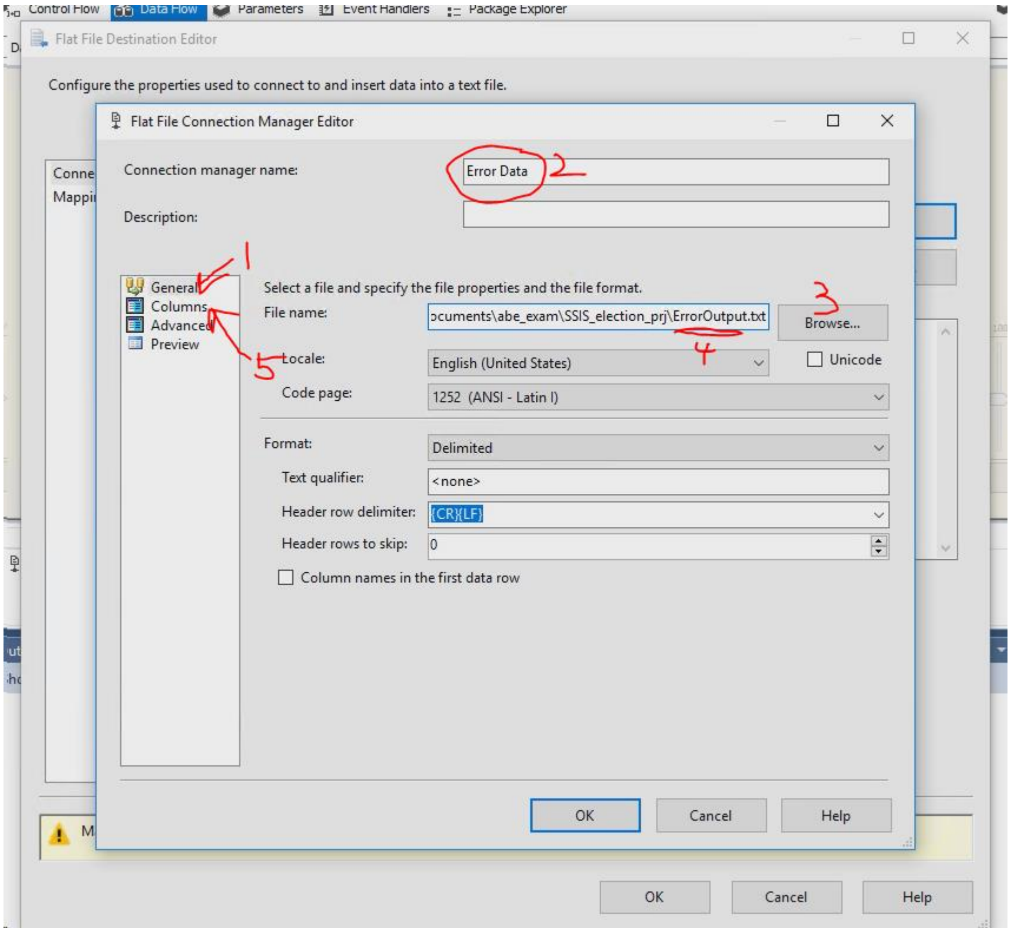
○ Fixed width with row delimiters

The columns are defined by fixed widths. An extra column, delimited by the new line characters, is added to define row delimiters.

○ Ragged right

The columns are defined by fixed widths, except the last one which is delimited by the new line character.

Map the columns on the Mappings page.



Flat File Destination Editor

Configure the properties used to connect to and insert data into a text file.

Flat File Connection Manager Editor

Connection manager name: Error Data

Description:

Select a file and specify the file properties and the file format.

File name: ɔcuments\abe_exam\SSIS_election_prj\ErrorOutput.txt

Locale: English (United States) □ Unicode

Code page: 1252 (ANSI - Latin I)

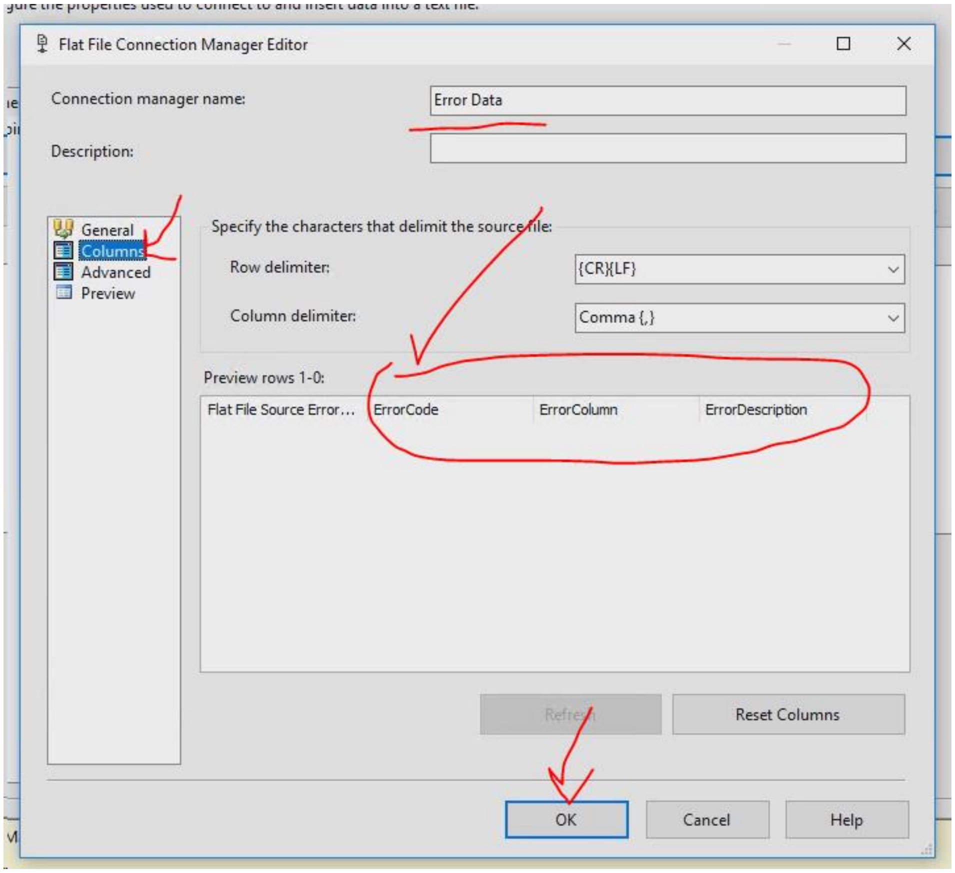
Format: Delimited

Text qualifier: <none>

Header row delimiter: {CR}{LF}

Header rows to skip: 0

□ Column names in the first data row



gure the properties used to connect to and insert data into a text file.

Connection manager name: Error Data

Description:

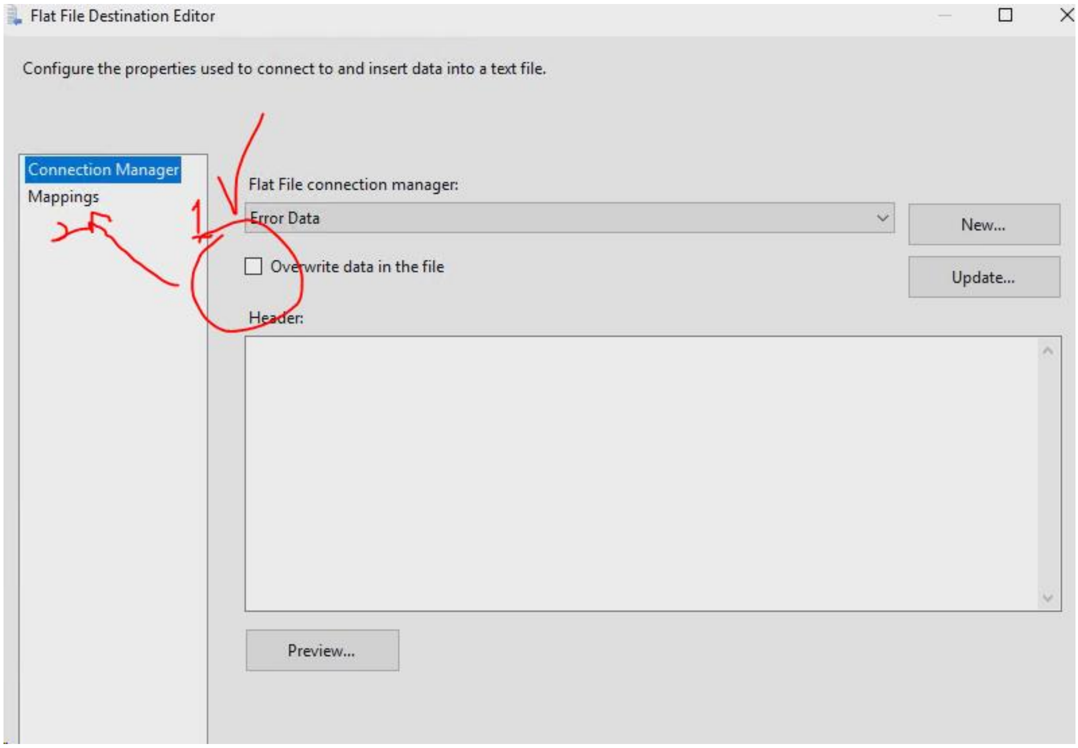
Specify the characters that delimit the source file:

Row delimiter: {CR}{LF}

Column delimiter: Comma {,}

Preview rows 1-0:

Flat File Source Error...	ErrorCode	ErrorColumn	ErrorDescription

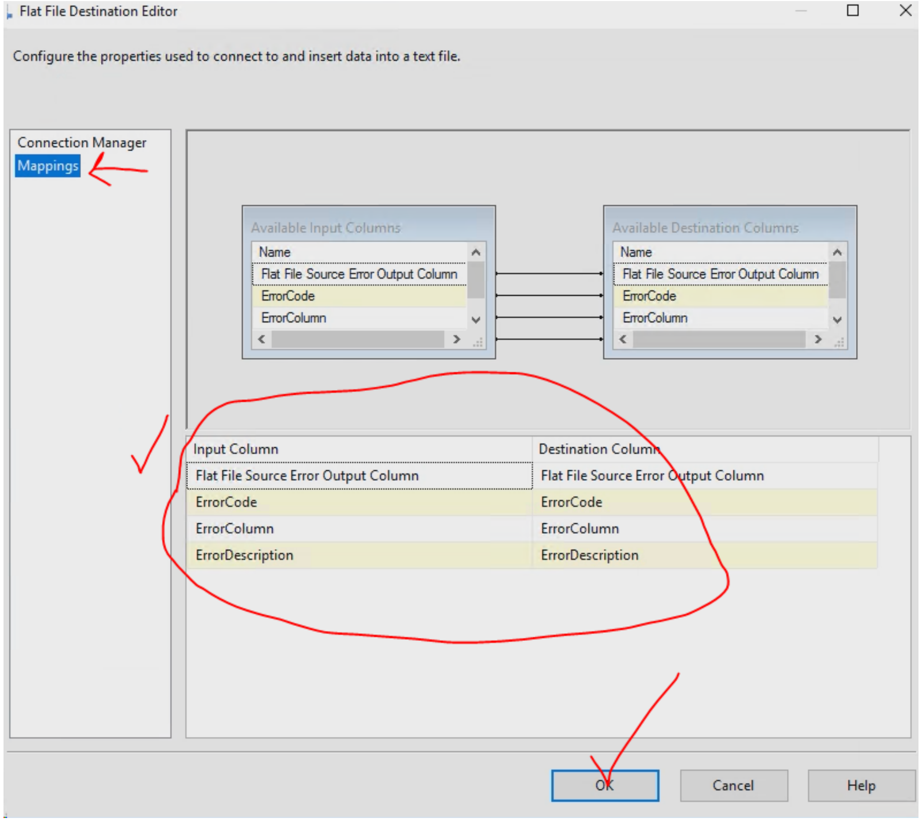


Configure the properties used to connect to and insert data into a text file.

Flat File connection manager: Error Data

□ Overwrite data in the file

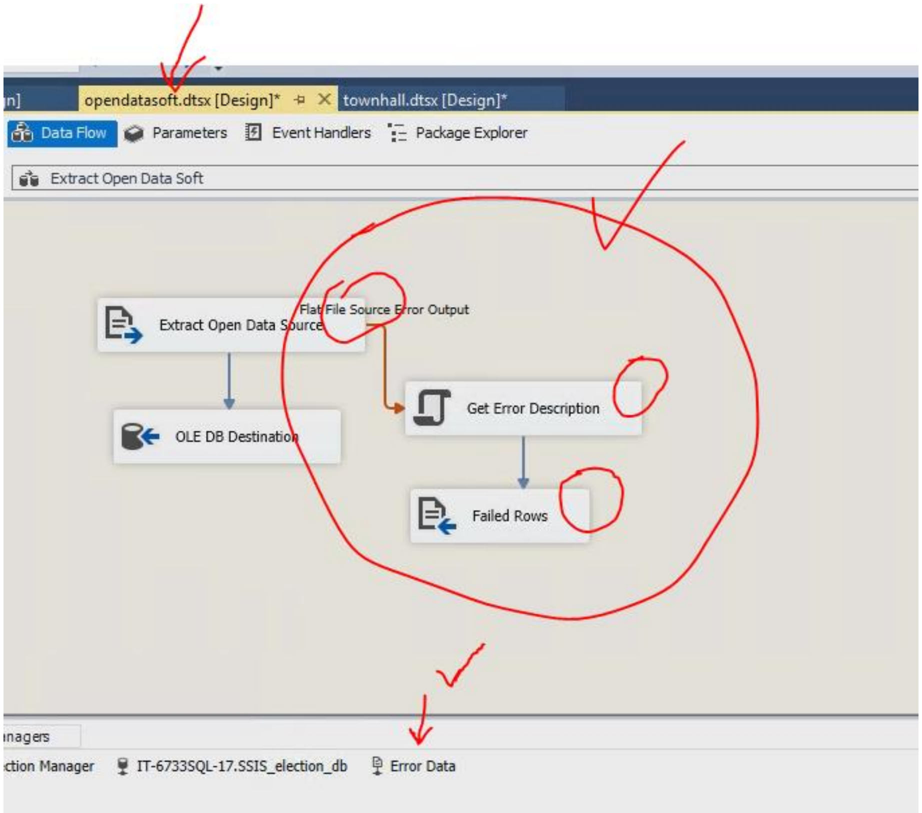
Header:



Flat File Destination Editor

Configure the properties used to connect to and insert data into a text file.

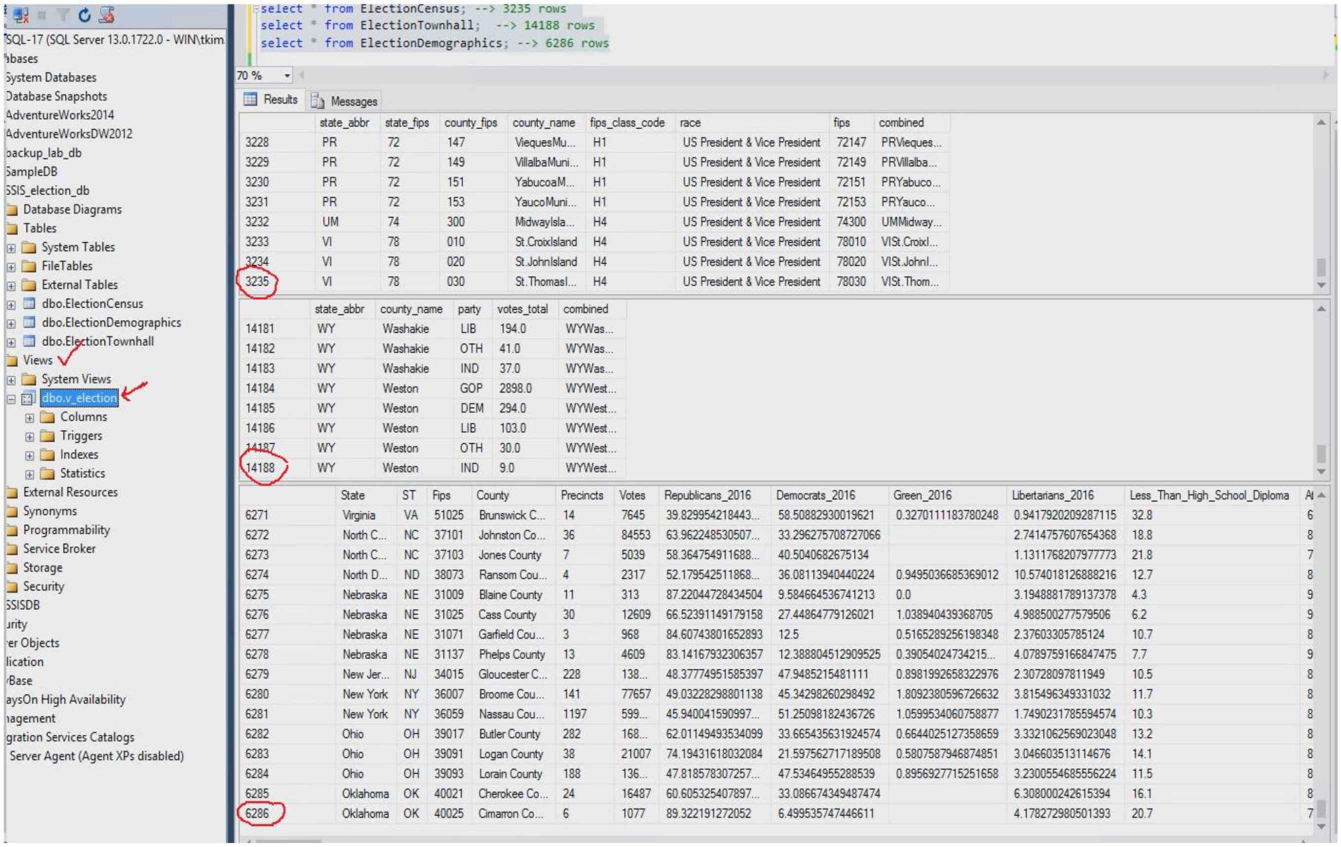
Input Column	Destination Column
Flat File Source Error Output Column	Flat File Source Error Output Column
ErrorCode	ErrorCode
ErrorColumn	ErrorColumn
ErrorDescription	ErrorDescription



43. ETL completed. Next test with SQL queries.

IV. Screenshots and SQL code (source code attached as separate file)

```
-- How many rows are there in each table?
select * from ElectionCensus;  -->3235 rows
select * from ElectionTownhall;  --> 14188 rows
select * from ElectionDemographics; --> 6286 rows

-- Create a view of ElectionCensus and ElectionTownhall
create view v_election as
select t.state_abbr
,t.county_name
,t.party
,t.votes_total
,s.fips
,s.state_fips
,s.county_fips
,s.fips_class_code
,s.race
from ElectionTownhall t
left outer join ElectionCensus s
on t.combined = s.combined;
```



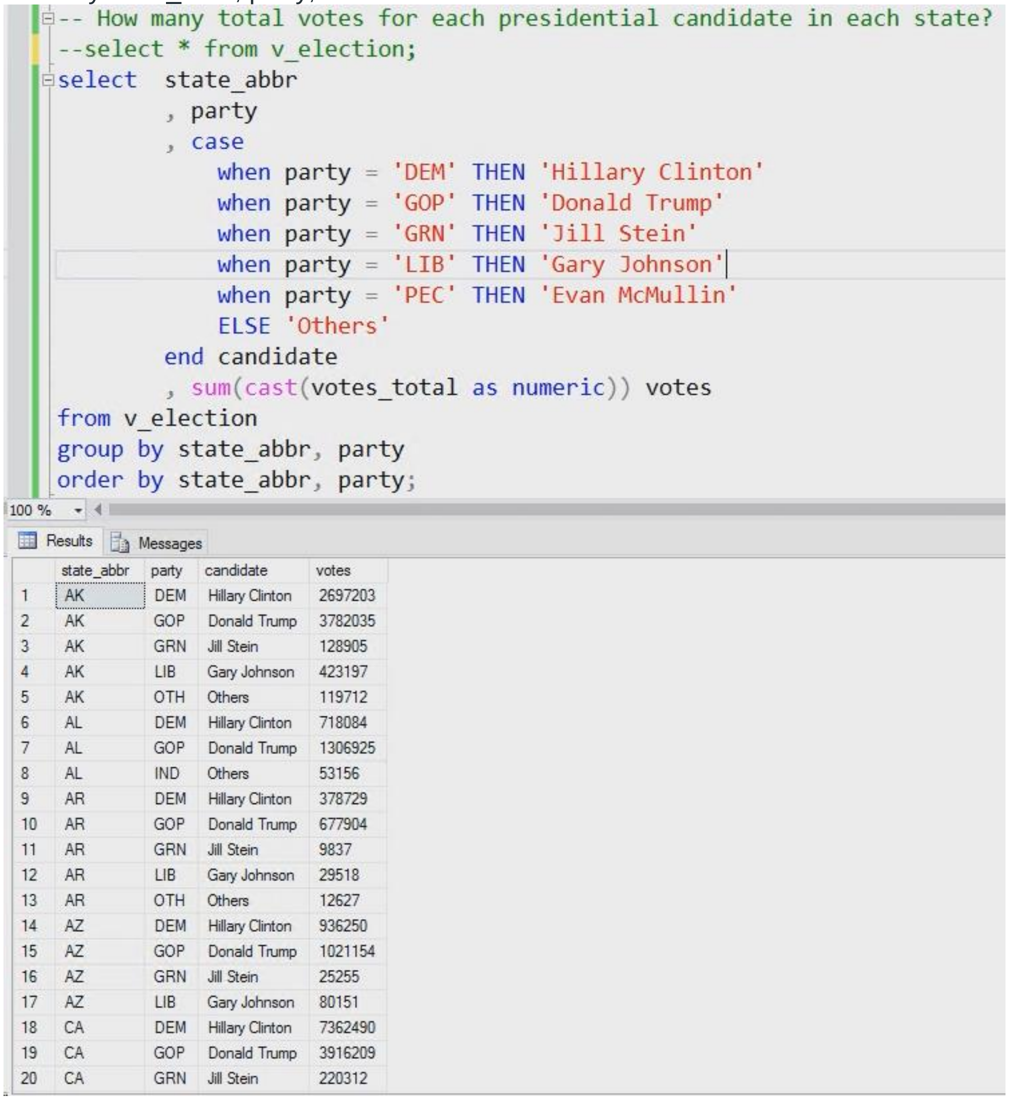
```
select * from ElectionCensus; --> 3235 rows
select * from ElectionTownhall;  --> 14188 rows
select * from ElectionDemographics; --> 6286 rows
```

	state_abbr	state_fips	county_fips	county_name	fips_class_code	race	fips	combined
3228	PR	72	147	ViequesMu...	H1	US President & Vice President	72147	PRVieques...
3229	PR	72	149	VillalbaMuni...	H1	US President & Vice President	72149	PRVillalba...
3230	PR	72	151	YabucoaM...	H1	US President & Vice President	72151	PRYabuco...
3231	PR	72	153	YaucoMuni...	H1	US President & Vice President	72153	PRYauco...
3232	UM	74	300	MidwayIsla...	H4	US President & Vice President	74300	UMMidway...
3233	VI	78	010	St.CroixIsland	H4	US President & Vice President	78010	VISt.CroixI...
3234	VI	78	020	St.JohnIsland	H4	US President & Vice President	78020	VISt.JohnI...
3235	VI	78	030	St.ThomasI...	H4	US President & Vice President	78030	VISt.Thom...

	state_abbr	county_name	party	votes_total	combined
14181	WY	Washakie	LIB	194.0	WYWas...
14182	WY	Washakie	OTH	41.0	WYWas...
14183	WY	Washakie	IND	37.0	WYWas...
14184	WY	Weston	GOP	2898.0	WYWest...
14185	WY	Weston	DEM	294.0	WYWest...
14186	WY	Weston	LIB	103.0	WYWest...
14187	WY	Weston	OTH	30.0	WYWest...
14188	WY	Weston	IND	9.0	WYWest...

	State	ST	Fips	County	Precincts	Votes	Republicans_2016	Democrats_2016	Green_2016	Libertarians_2016	Less_Than_High_School_Diploma	At
6271	Virginia	VA	51025	Brunswick C...	14	7645	39.829954218443...	58.50882930019621	0.3270111183780248	0.9417920209287115	32.8	6
6272	North C...	NC	37101	Johnston Co...	36	84553	63.962248530507...	33.296275708727066		2.7414757607654368	18.8	8
6273	North C...	NC	37103	Jones County	7	5039	58.364754911688...	40.5040682675134		1.1311768207977773	21.8	7
6274	North D...	ND	38073	Ransom Cou...	4	2317	52.179542511868...	36.08113940440224	0.9495036685369012	10.574018126888216	12.7	8
6275	Nebraska	NE	31009	Blaine County	11	313	87.22044728434504	9.584664536741213	0.0	3.1948881789137378	4.3	9
6276	Nebraska	NE	31025	Cass County	30	12609	66.52391149179158	27.44864779126021	1.038940439368705	4.988500277579506	6.2	9
6277	Nebraska	NE	31071	Garfield Cou...	3	968	84.60743801652893	12.5	0.5165289256198348	2.37603305785124	10.7	8
6278	Nebraska	NE	31137	Phelps County	13	4609	83.14167932306357	12.388804512909525	0.39054024734215...	4.0789759166847475	7.7	9
6279	New Jer...	NJ	34015	Gloucester C...	228	138...	48.37774951585397	47.9485215481111	0.8981992658322976	2.30728097811949	10.5	8
6280	New York	NY	36007	Broome Cou...	141	77657	49.03228298801138	45.34298260298492	1.8092380596726632	3.815496349331032	11.7	8
6281	New York	NY	36059	Nassau Cou...	1197	599...	45.940041590997...	51.25098182436726	1.0599534060758877	1.7490231785594574	10.3	8
6282	Ohio	OH	39017	Butler County	282	168...	62.01149493534099	33.665435631924574	0.6644025127358659	3.3321062569023048	13.2	8
6283	Ohio	OH	39091	Logan County	38	21007	74.19431618032084	21.597562717189508	0.5807587946874851	3.046603513114676	14.1	8
6284	Ohio	OH	39093	Lorain County	188	136...	47.818578307257...	47.53464955288539	0.8956927715251658	3.2300554685556224	11.5	8
6285	Oklahoma	OK	40021	Cherokee Co...	24	16487	60.605325407897...	33.086674349487474		6.308000242615394	16.1	8
6286	Oklahoma	OK	40025	Cimarron Co...	6	1077	89.322191272052	6.499535747446611		4.178272980501393	20.7	7

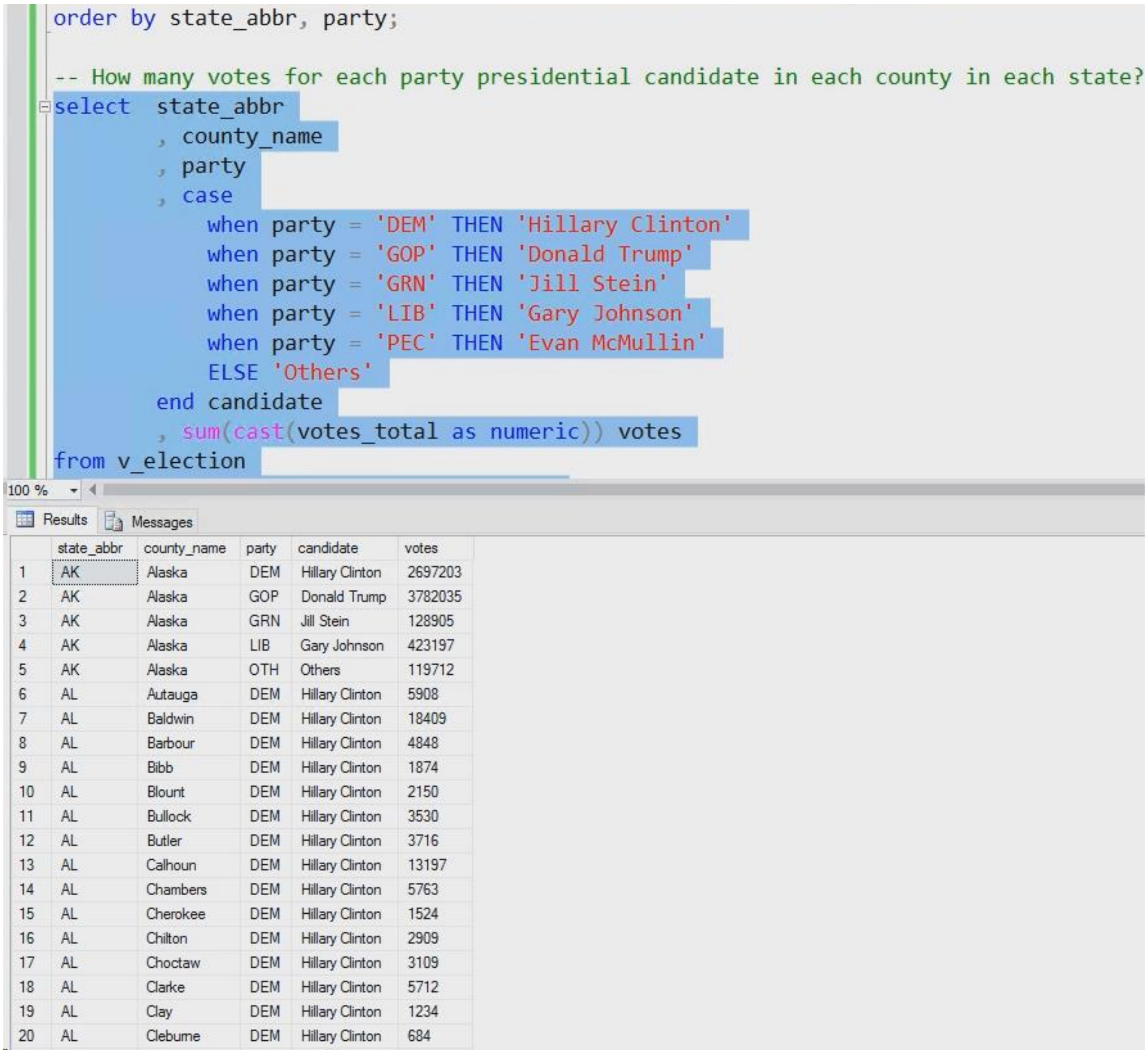
```
-- How many total votes for each presidential candidate in each state?
select * from v_election;
select state_abbr
, party
, case
when party = 'DEM' THEN 'Hillary Clinton'
when party = 'GOP' THEN 'Donald Trump'
when party = 'GRN' THEN 'Jill Stein'
when party = 'LIB' THEN 'Gary Johnson'
when party = 'PEC' THEN 'Evan McMullin'
ELSE 'Others'
end candidate
, sum(cast(votes_total as numeric)) votes
from v_election
group by state_abbr, party
order by state_abbr, party;
```



```
-- How many total votes for each presidential candidate in each state?
--select * from v_election;
select  state_abbr
        , party
        , case
            when party = 'DEM' THEN 'Hillary Clinton'
            when party = 'GOP' THEN 'Donald Trump'
            when party = 'GRN' THEN 'Jill Stein'
            when party = 'LIB' THEN 'Gary Johnson'
            when party = 'PEC' THEN 'Evan McMullin'
            ELSE 'Others'
        end candidate
        , sum(cast(votes_total as numeric)) votes
from v_election
group by state_abbr, party
order by state_abbr, party;
```

	state_abbr	party	candidate	votes
1	AK	DEM	Hillary Clinton	2697203
2	AK	GOP	Donald Trump	3782035
3	AK	GRN	Jill Stein	128905
4	AK	LIB	Gary Johnson	423197
5	AK	OTH	Others	119712
6	AL	DEM	Hillary Clinton	718084
7	AL	GOP	Donald Trump	1306925
8	AL	IND	Others	53156
9	AR	DEM	Hillary Clinton	378729
10	AR	GOP	Donald Trump	677904
11	AR	GRN	Jill Stein	9837
12	AR	LIB	Gary Johnson	29518
13	AR	OTH	Others	12627
14	AZ	DEM	Hillary Clinton	936250
15	AZ	GOP	Donald Trump	1021154
16	AZ	GRN	Jill Stein	25255
17	AZ	LIB	Gary Johnson	80151
18	CA	DEM	Hillary Clinton	7362490
19	CA	GOP	Donald Trump	3916209
20	CA	GRN	Jill Stein	220312

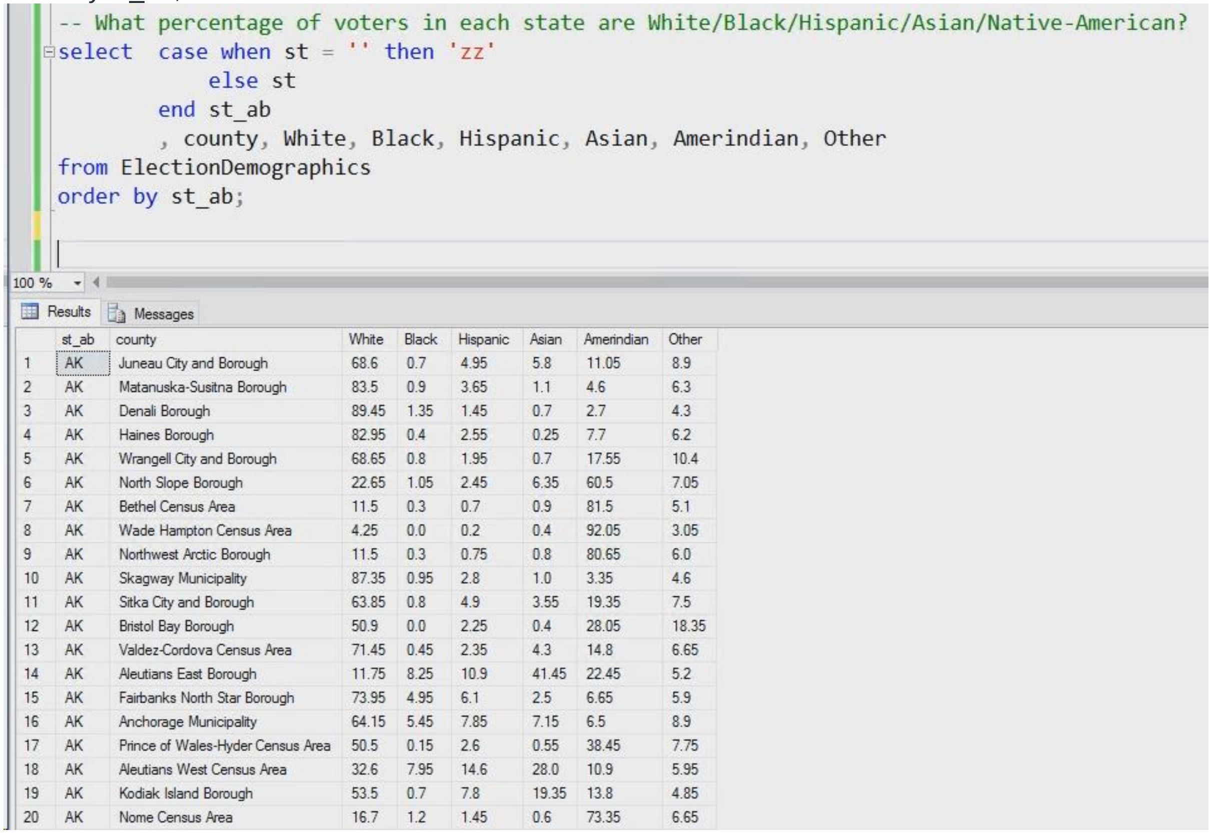
```
-- How many votes for each party presidential candidate in each county in each state?
select state_abbr
, county_name
, party
, case
when party = 'DEM' THEN 'Hillary Clinton'
when party = 'GOP' THEN 'Donald Trump'
when party = 'GRN' THEN 'Jill Stein'
when party = 'LIB' THEN 'Gary Johnson'
when party = 'PEC' THEN 'Evan McMullin'
ELSE 'Others'
end candidate
, sum(cast(votes_total as numeric)) votes
from v_election
group by state_abbr, party, county_name
order by state_abbr, party, county_name;
```



```
order by state_abbr, party;

-- How many votes for each party presidential candidate in each county in each state?
select  state_abbr
        , county_name
        , party
        , case
            when party = 'DEM' THEN 'Hillary Clinton'
            when party = 'GOP' THEN 'Donald Trump'
            when party = 'GRN' THEN 'Jill Stein'
            when party = 'LIB' THEN 'Gary Johnson'
            when party = 'PEC' THEN 'Evan McMullin'
            ELSE 'Others'
        end candidate
        , sum(cast(votes_total as numeric)) votes
from v_election
```

	state_abbr	county_name	party	candidate	votes
1	AK	Alaska	DEM	Hillary Clinton	2697203
2	AK	Alaska	GOP	Donald Trump	3782035
3	AK	Alaska	GRN	Jill Stein	128905
4	AK	Alaska	LIB	Gary Johnson	423197
5	AK	Alaska	OTH	Others	119712
6	AL	Autauga	DEM	Hillary Clinton	5908
7	AL	Baldwin	DEM	Hillary Clinton	18409
8	AL	Barbour	DEM	Hillary Clinton	4848
9	AL	Bibb	DEM	Hillary Clinton	1874
10	AL	Blount	DEM	Hillary Clinton	2150
11	AL	Bullock	DEM	Hillary Clinton	3530
12	AL	Butler	DEM	Hillary Clinton	3716
13	AL	Calhoun	DEM	Hillary Clinton	13197
14	AL	Chambers	DEM	Hillary Clinton	5763
15	AL	Cherokee	DEM	Hillary Clinton	1524
16	AL	Chilton	DEM	Hillary Clinton	2909
17	AL	Choctaw	DEM	Hillary Clinton	3109
18	AL	Clarke	DEM	Hillary Clinton	5712
19	AL	Clay	DEM	Hillary Clinton	1234
20	AL	Clebume	DEM	Hillary Clinton	684

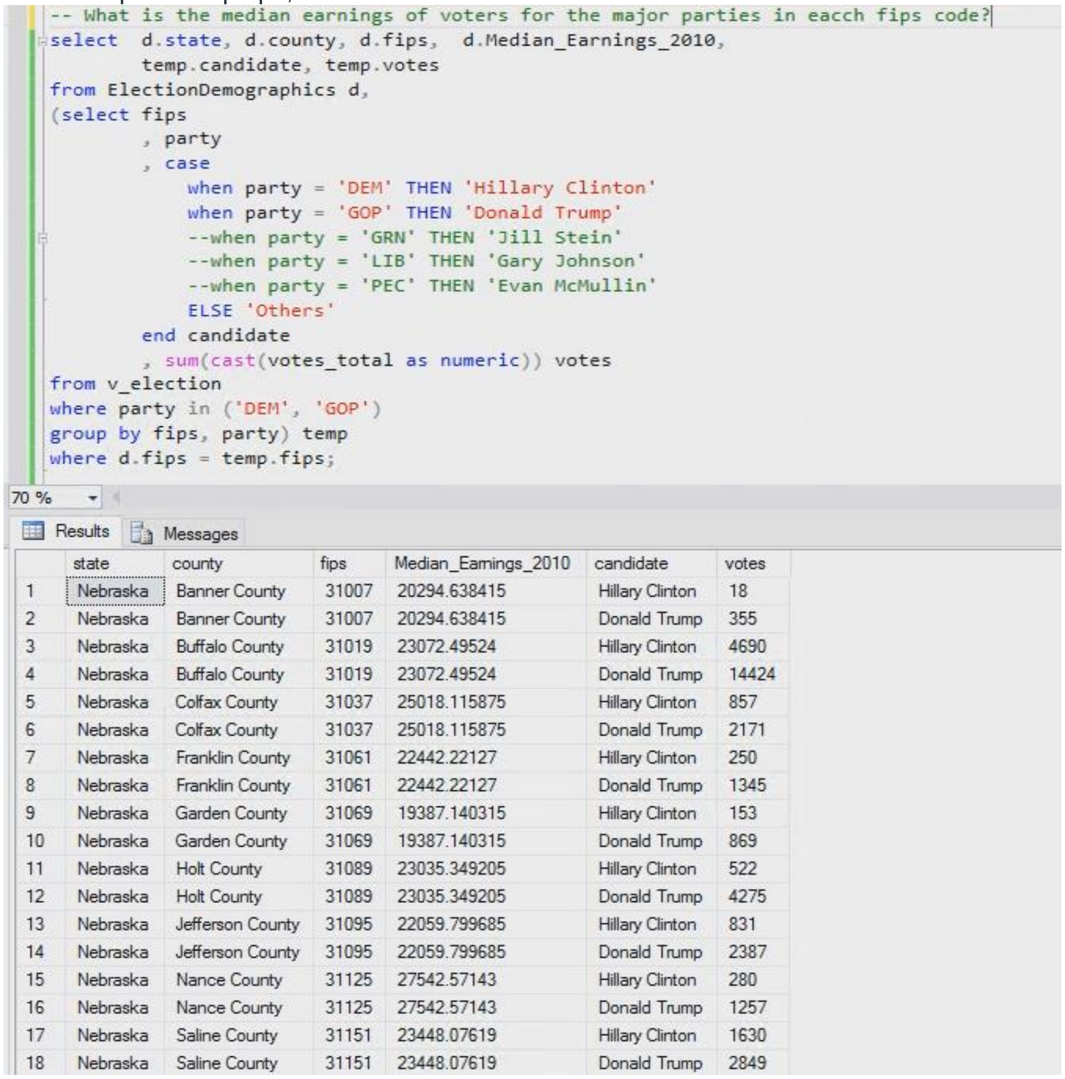
```
-- What percentage of voters in each state are White/Black/Hispanic/Asian/Native-American?
select case when st = '' then 'zz'
else st
end st_ab
, county, White, Black, Hispanic, Asian, Amerindian, Other
from ElectionDemographics
order by st_ab;
```



```
-- What percentage of voters in each state are White/Black/Hispanic/Asian/Native-American?
select  case when st = '' then 'zz'
            else st
        end st_ab
        , county, White, Black, Hispanic, Asian, Amerindian, Other
from ElectionDemographics
order by st_ab;
```

	st_ab	county	White	Black	Hispanic	Asian	Amerindian	Other
1	AK	Juneau City and Borough	68.6	0.7	4.95	5.8	11.05	8.9
2	AK	Matanuska-Susitna Borough	83.5	0.9	3.65	1.1	4.6	6.3
3	AK	Denali Borough	89.45	1.35	1.45	0.7	2.7	4.3
4	AK	Haines Borough	82.95	0.4	2.55	0.25	7.7	6.2
5	AK	Wrangell City and Borough	68.65	0.8	1.95	0.7	17.55	10.4
6	AK	North Slope Borough	22.65	1.05	2.45	6.35	60.5	7.05
7	AK	Bethel Census Area	11.5	0.3	0.7	0.9	81.5	5.1
8	AK	Wade Hampton Census Area	4.25	0.0	0.2	0.4	92.05	3.05
9	AK	Northwest Arctic Borough	11.5	0.3	0.75	0.8	80.65	6.0
10	AK	Skagway Municipality	87.35	0.95	2.8	1.0	3.35	4.6
11	AK	Sitka City and Borough	63.85	0.8	4.9	3.55	19.35	7.5
12	AK	Bristol Bay Borough	50.9	0.0	2.25	0.4	28.05	18.35
13	AK	Valdez-Cordova Census Area	71.45	0.45	2.35	4.3	14.8	6.65
14	AK	Aleutians East Borough	11.75	8.25	10.9	41.45	22.45	5.2
15	AK	Fairbanks North Star Borough	73.95	4.95	6.1	2.5	6.65	5.9
16	AK	Anchorage Municipality	64.15	5.45	7.85	7.15	6.5	8.9
17	AK	Prince of Wales-Hyder Census Area	50.5	0.15	2.6	0.55	38.45	7.75
18	AK	Aleutians West Census Area	32.6	7.95	14.6	28.0	10.9	5.95
19	AK	Kodiak Island Borough	53.5	0.7	7.8	19.35	13.8	4.85
20	AK	Nome Census Area	16.7	1.2	1.45	0.6	73.35	6.65

```
-- What is the median earnings of voters for the major parties in each fips code?
select d.state, d.county, d.fips,  d.Median_Earnings_2010,
temp.candidate, temp.votes
from ElectionDemographics d,
(select fips
, party
, case
when party = 'DEM' THEN 'Hillary Clinton'
when party = 'GOP' THEN 'Donald Trump'
--when party = 'GRN' THEN 'Jill Stein'
--when party = 'LIB' THEN 'Gary Johnson'
--when party = 'PEC' THEN 'Evan McMullin'
ELSE 'Others'
end candidate
, sum(cast(votes_total as numeric)) votes
from v_election
where party in ('DEM', 'GOP')
group by fips, party) temp
where d.fips = temp.fips;
```



```
-- What is the median earnings of voters for the major parties in eacch fips code?
select  d.state, d.county, d.fips,  d.Median_Earnings_2010,
        temp.candidate, temp.votes
from ElectionDemographics d,
(select fips
        , party
        , case
            when party = 'DEM' THEN 'Hillary Clinton'
            when party = 'GOP' THEN 'Donald Trump'
            --when party = 'GRN' THEN 'Jill Stein'
            --when party = 'LIB' THEN 'Gary Johnson'
            --when party = 'PEC' THEN 'Evan McMullin'
            ELSE 'Others'
        end candidate
        , sum(cast(votes_total as numeric)) votes
from v_election
where party in ('DEM', 'GOP')
group by fips, party) temp
where d.fips = temp.fips;
```

	state	county	fips	Median_Earnings_2010	candidate	votes
1	Nebraska	Banner County	31007	20294.638415	Hillary Clinton	18
2	Nebraska	Banner County	31007	20294.638415	Donald Trump	355
3	Nebraska	Buffalo County	31019	23072.49524	Hillary Clinton	4690
4	Nebraska	Buffalo County	31019	23072.49524	Donald Trump	14424
5	Nebraska	Colfax County	31037	25018.115875	Hillary Clinton	857
6	Nebraska	Colfax County	31037	25018.115875	Donald Trump	2171
7	Nebraska	Franklin County	31061	22442.22127	Hillary Clinton	250
8	Nebraska	Franklin County	31061	22442.22127	Donald Trump	1345
9	Nebraska	Garden County	31069	19387.140315	Hillary Clinton	153
10	Nebraska	Garden County	31069	19387.140315	Donald Trump	869
11	Nebraska	Holt County	31089	23035.349205	Hillary Clinton	522
12	Nebraska	Holt County	31089	23035.349205	Donald Trump	4275
13	Nebraska	Jefferson County	31095	22059.799685	Hillary Clinton	831
14	Nebraska	Jefferson County	31095	22059.799685	Donald Trump	2387
15	Nebraska	Nance County	31125	27542.57143	Hillary Clinton	280
16	Nebraska	Nance County	31125	27542.57143	Donald Trump	1257
17	Nebraska	Saline County	31151	23448.07619	Hillary Clinton	1630
18	Nebraska	Saline County	31151	23448.07619	Donald Trump	2849

V. Detailed description of each group member contribution

Member	Task	Description	Estimated time (min)
Abraham Kim	Research potential topics	Google search for topics with sufficient number of separate sources	180
	Discuss/Evaluate options	Narrow choices based on relevance to project goals and select topic	100
	Downloaded two data files	Download (Townhall source and US Census source), scrape and transform Townhall source data from html to csv using Python	120
	Practice loading into SSIS	Test load xls and txt file into SQL Server SSIS	240
	Data cleansing	Data cleansing of four sources (using Python)	600
	Build & deploy ETL	Fix errors & debug during loading of files	600
		Run to test successful load, capture screenshots	80
	Create report	Support creation of report	60
Karis Kim	Research potential topics	Google search for topics and types of data openly available	60
	Discuss/Evaluate options	Narrow choices based on relevance to project goals and select topic	100
	Find, download two additional data sources	OpenDataSoft source and Data.gov source as csv files	120
	Prepare report template	Study project deliverables, create report template	30
	Review ETL procedures from Lab	Reference Lab 9 on SSIS procedures	30
	Build & deploy ETL	Support data cleansing	30
		Create DB in SQL Server and create project in SS Data Tool	30
		Load data sources in SQL Server Data Tool 2015	600
		Capture screenshots	300
		Create queries to test	30
	Create report	Create report	360