File: DSC_630_Course_Project_Preliminary_Analysis_Final.ipynb
Names: Corinne Medeiros, Amy Nestingen
Date: 11/12/20
Course: DSC 630 Predictive Analytics
Desc: Course Project - Preliminary Analysis with updated data
Usage: Program cleans data, generates exploratory visualizations, and saves cleaned data to a csv file.

```python
In [1]: import matplotlib
        import matplotlib.pyplot as plt

        import pandas as pd
        import numpy as np

        import matplotlib.dates as mdates
        from matplotlib.dates import DateFormatter
```

DSC 630 Course Project - Preliminary Analysis with updated data

Predicting Travel Trends in the United States 2019 - 2020

Data source: https://catalog.data.gov/dataset/trips-by-distance (https://catalog.data.gov/dataset/trips-by-distance)

Loading and Previewing Data

```python
In [2]: # Loading data into a Pandas DataFrame
        trip_data = pd.read_csv("Trips_by_Distance.csv")
```

```python
In [3]: # Checking dimensions
        print(trip_data.shape)

        (2117622, 19)
```

```python
In [4]: # Previewing data
        trip_data.head(5)
```

Out[4]:

	Level	Date	State FIPS	State Postal Code	County FIPS	County Name	Population Staying at Home	Population Not Staying at Home	Number of Trips	Number of Trips <1	Number of Trips 1-3	Number of Trips 3-5	Number of Trips 5-10	Number of Trips 10-25	Number of Trips 25-50
0	County	2019/01/01	29.0	MO	29171.0	Putnam County	1155.0	3587.0	12429.0	2807.0	3642.0	1272.0	1240.0	1953.0	1058.0
1	County	2019/01/01	2.0	AK	2164.0	Lake and Peninsula Borough	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	County	2019/01/01	1.0	AL	1001.0	Autauga County	9624.0	45807.0	132004.0	27097.0	35263.0	18315.0	18633.0	22963.0	5149.0
3	County	2019/01/01	1.0	AL	1003.0	Baldwin County	44415.0	172941.0	534520.0	120752.0	142931.0	68235.0	87430.0	78045.0	24495.0
4	County	2019/01/01	1.0	AL	1005.0	Barbour County	4782.0	20023.0	67658.0	15524.0	16677.0	10550.0	11674.0	6416.0	3686.0

Displaying Data Summaries

```python
In [5]: # Displaying summary information for string columns

        print("String Data:\n")
        print(trip_data.describe(include=['O']))

        String Data:

                 Level        Date State Postal Code       County Name
        count  2117622     2117622            2116959           2083146
        unique       3         663                 51              1877
        top     County  2019/11/24                 TX  Washington County
        freq   2083146        3194             169065             19890
```

```python
# Displaying summary information for numeric columns without scientific notation

with pd.option_context('float_format', '{:f}'.format):
    print("Numeric Data:\n\n", trip_data.describe())
```

Numeric Data:

	State FIPS	County FIPS	Population Staying at Home \
count	2116959.000000	2083146.000000	2094677.000000
mean	30.259004	30383.649268	66586.202567
std	15.151528	15160.098945	1291456.301907
min	1.000000	1001.000000	8.000000
25%	18.000000	18177.000000	2134.000000
50%	29.000000	29176.000000	5045.000000
75%	45.000000	45081.000000	14828.000000
max	56.000000	56045.000000	110211784.000000

	Population Not Staying at Home	Number of Trips	Number of Trips <1 \
count	2094677.000000	2094677.000000	2094677.000000
mean	243536.388818	1140437.857211	278926.770685
std	4669586.944696	22182683.988919	5416049.593708
min	-38.000000	220.000000	0.000000
25%	9303.000000	42765.000000	9099.000000
50%	21869.000000	102799.000000	22174.000000
75%	59629.000000	283165.000000	62541.000000
max	273739951.000000	1569052522.000000	422700184.000000

	Number of Trips 1-3	Number of Trips 3-5	Number of Trips 5-10 \
count	2094677.000000	2094677.000000	2094677.000000
mean	286260.461260	139756.301626	176731.594259
std	5578488.538858	2724288.872000	3452786.729973
min	0.000000	-108.000000	0.000000
25%	9627.000000	4066.000000	5572.000000
50%	25519.000000	11690.000000	14393.000000
75%	73530.000000	35286.000000	42045.000000
max	405130498.000000	198018442.000000	252611815.000000

	Number of Trips 10-25	Number of Trips 25-50	Number of Trips 50-100 \
count	2094677.000000	2094677.000000	2094677.000000
mean	174372.367290	56125.496253	17765.654940
std	3417782.033598	1089528.996507	342504.165421
min	0.000000	0.000000	0.000000
25%	6802.000000	3359.000000	1167.000000
50%	16262.000000	7249.000000	2571.000000
75%	43141.000000	17185.000000	5951.000000
max	256509626.000000	76367324.000000	25539735.000000

	Number of Trips 100-250	Number of Trips 250-500	Number of Trips >=500
count	2094677.000000	2094677.000000	2094677.000000
mean	7275.363191	1679.449165	1544.398543
std	142671.799201	33080.770917	34392.715818
min	0.000000	0.000000	0.000000
25%	370.000000	56.000000	19.000000
50%	873.000000	167.000000	68.000000
75%	2230.000000	473.000000	262.000000
max	14476977.000000	3651375.000000	5003062.000000

```
In [7]:  # List unique values in the County Name column
         counties_list = trip_data['County Name'].unique().tolist()

         print(counties_list)
```

```
['Putnam County', 'Lake and Peninsula Borough', 'Autauga County', 'Baldwin County', 'Barbour County', 'Bibb C
ounty', 'Blount County', 'Bullock County', 'Butler County', 'Calhoun County', 'Chambers County', 'Cherokee Co
unty', 'Chilton County', 'Choctaw County', 'Clarke County', 'Clay County', 'Cleburne County', 'Coffee Count
y', 'Colbert County', 'Conecuh County', 'Coosa County', 'Covington County', 'Crenshaw County', 'Cullman Count
y', 'Dale County', 'Dallas County', 'DeKalb County', 'Elmore County', 'Escambia County', 'Etowah County', 'Fa
yette County', 'Franklin County', 'Geneva County', 'Greene County', 'Hale County', 'Matanuska-Susitna Boroug
h', 'Nome Census Area', 'North Slope Borough', 'Northwest Arctic Borough', 'Petersburg Borough', 'Prince of W
ales-Hyder Census Area', 'Sitka City and Borough', 'Skagway Municipality', 'Southeast Fairbanks Census Area',
'Valdez-Cordova Census Area', 'Wrangell City and Borough', 'Yakutat City and Borough', 'Yukon-Koyukuk Census
Area', 'Apache County', 'Cochise County', 'Coconino County', 'Gila County', 'Graham County', 'Greenlee Count
y', 'La Paz County', 'Maricopa County', 'Mohave County', 'Navajo County', 'Pima County', 'Pinal County', 'San
ta Cruz County', 'Yavapai County', 'Yuma County', 'Arkansas County', 'Ashley County', 'Baxter County', 'Bento
n County', 'Boone County', 'Bradley County', 'Carroll County', 'Chicot County', 'Clark County', 'Cleveland Co
unty', 'Columbia County', 'Conway County', 'Craighead County', 'Crawford County', 'Crittenden County', 'Cross
County', 'Desha County', 'Van Buren County', 'Washington County', 'White County', 'Woodruff County', 'Yell Co
unty', 'Alameda County', 'Alpine County', 'Amador County', 'Butte County', 'Calaveras County', 'Colusa Count
y', 'Contra Costa County', 'Del Norte County', 'El Dorado County', 'Fresno County', 'Glenn County', 'Humboldt
County', 'Imperial County', 'Inyo County', 'Kern County', 'Kings County', 'Lake County', 'Lassen County', 'Lo
s Angeles County', 'Madera County', 'Marin County', 'Mariposa County', 'Mendocino County', 'Merced County',
'Modoc County', 'Mono County', 'Monterey County', 'Napa County', 'Nevada County', 'Orange County', 'Placer Co
unty', 'Plumas County', 'Riverside County', 'Sacramento County', 'San Benito County', 'San Bernardino Count
y', 'San Diego County', 'San Francisco County', 'San Joaquin County', 'San Luis Obispo County', 'San Mateo Co
unty', 'Santa Barbara County', 'Santa Clara County', 'La Plata County', 'Larimer County', 'Las Animas Count
y', 'Lincoln County', 'Logan County', 'Mesa County', 'Mineral County', 'Moffat County', 'Montezuma County',
'Montrose County', 'Morgan County', 'Otero County', 'Ouray County', 'Park County', 'Phillips County', 'Pitkin
County', 'Prowers County', 'Pueblo County', 'Rio Blanco County', 'Rio Grande County', 'Routt County', 'Saguac
he County', 'San Juan County', 'San Miguel County', 'Sedgwick County', 'Summit County', 'Teller County', 'Wel
d County', 'Fairfield County', 'Hartford County', 'Litchfield County', 'Middlesex County', 'New Haven Count
y', 'New London County', 'Tolland County', 'Windham County', 'Kent County', 'New Castle County', 'Sussex Coun
ty', 'District of Columbia', 'Alachua County', 'Baker County', 'Bay County', 'Bradford County', 'Brevard Coun
ty', 'Broward County', 'Santa Rosa County', 'Sarasota County', 'Seminole County', 'Sumter County', 'Suwannee
County', 'Taylor County', 'Union County', 'Volusia County', 'Wakulla County', 'Walton County', 'Appling Count
y', 'Atkinson County', 'Bacon County', 'Banks County', 'Barrow County', 'Bartow County', 'Ben Hill County',
'Berrien County', 'Bleckley County', 'Brantley County', 'Brooks County', 'Bryan County', 'Bulloch County', 'B
urke County', 'Butts County', 'Camden County', 'Candler County', 'Catoosa County', 'Charlton County', 'Chatha
m County', 'Chattahoochee County', 'Chattooga County', 'Clayton County', 'Clinch County', 'Cobb County', 'Col
quitt County', 'Cook County', 'Coweta County', 'Crisp County', 'Dade County', 'Dawson County', 'Decatur Count
y', 'Dodge County', 'Dooly County', 'Dougherty County', 'Douglas County', 'Early County', 'Echols County', 'E
ffingham County', 'Elbert County', 'Emanuel County', 'Evans County', 'Fannin County', 'Floyd County', 'Forsyt
h County', 'Fulton County', 'Gilmer County', 'Glascock County', 'Glynn County', 'Gordon County', 'Grady Count
y', 'Gwinnett County', 'Habersham County', 'Hall County', 'Hancock County', 'Haralson County', 'Harris Count
y', 'Hart County', 'Heard County', 'Henry County', 'Houston County', 'Irwin County', 'Jackson County', 'Jaspe
r County', 'Jeff Davis County', 'Jefferson County', 'Jenkins County', 'Johnson County', 'Jones County', 'Lama
r County', 'Lanier County', 'Laurens County', 'Tift County', 'Toombs County', 'Towns County', 'Treutlen Count
y', 'Troup County', 'Turner County', 'Twiggs County', 'Upson County', 'Walker County', 'Ware County', 'Warren
County', 'Wayne County', 'Webster County', 'Wheeler County', 'Whitfield County', 'Wilcox County', 'Wilkes Cou
nty', 'Wilkinson County', 'Worth County', 'Hawaii County', 'Honolulu County', 'Kalawao County', 'Kauai Count
y', 'Maui County', 'Ada County', 'Adams County', 'Bannock County', 'Bear Lake County', 'Benewah County', 'Bin
gham County', 'Blaine County', 'Boise County', 'Bonner County', 'Bonneville County', 'Boundary County', 'Cama
s County', 'Canyon County', 'Caribou County', 'Cassia County', 'Clearwater County', 'Custer County', 'Fremont
County', 'Gallatin County', 'Grundy County', 'Hamilton County', 'Hardin County', 'Henderson County', 'Iroquoi
s County', 'Jersey County', 'Jo Daviess County', 'Kane County', 'Kankakee County', 'Kendall County', 'Knox Co
unty', 'LaSalle County', 'Lawrence County', 'Lee County', 'Livingston County', 'McDonough County', 'McHenry C
ounty', 'McLean County', 'Macon County', 'Macoupin County', 'Madison County', 'Marion County', 'Marshall Coun
ty', 'Mason County', 'Massac County', 'Menard County', 'Mercer County', 'Monroe County', 'Montgomery County',
'Moultrie County', 'Ogle County', 'Peoria County', 'Perry County', 'Piatt County', 'Pike County', 'Pope Count
y', 'Gibson County', 'Grant County', 'Harrison County', 'Hendricks County', 'Howard County', 'Huntington Coun
ty', 'Jay County', 'Jennings County', 'Kosciusko County', 'LaGrange County', 'LaPorte County', 'Martin Count
y', 'Miami County', 'Newton County', 'Noble County', 'Ohio County', 'Owen County', 'Parke County', 'Porter Co
unty', 'Posey County', 'Pulaski County', 'Randolph County', 'Ripley County', 'Rush County', 'St. Joseph Count
y', 'Scott County', 'Dickinson County', 'Dubuque County', 'Emmet County', 'Guthrie County', 'Ida County', 'Io
wa County', 'Keokuk County', 'Kossuth County', 'Linn County', 'Louisa County', 'Lucas County', 'Lyon County',
'Mahaska County', 'Mills County', 'Mitchell County', 'Monona County', 'Muscatine County', "O'Brien County",
'Osceola County', 'Page County', 'Palo Alto County', 'Plymouth County', 'Pocahontas County', 'Polk County',
'Pottawattamie County', 'Reno County', 'Republic County', 'Rice County', 'Riley County', 'Rooks County', 'Rus
sell County', 'Saline County', 'Seward County', 'Shawnee County', 'Sheridan County', 'Sherman County', 'Smith
County', 'Stafford County', 'Stanton County', 'Stevens County', 'Sumner County', 'Thomas County', 'Trego Coun
ty', 'Wabaunsee County', 'Wallace County', 'Wichita County', 'Wilson County', 'Woodson County', 'Wyandotte Co
unty', 'Adair County', 'Allen County', 'Anderson County', 'Ballard County', 'Barren County', 'Bath County',
'Bell County', 'Bourbon County', 'Boyd County', 'Boyle County', 'Bracken County', 'Breathitt County', 'Brecki
nridge County', 'Bullitt County', 'Caldwell County', 'Calloway County', 'Campbell County', 'Carlisle County',
'McCracken County', 'McCreary County', 'Magoffin County', 'Meade County', 'Menifee County', 'Metcalfe Count
y', 'Muhlenberg County', 'Nelson County', 'Nicholas County', 'Oldham County', 'Owsley County', 'Pendleton Cou
nty', 'Powell County', 'Robertson County', 'Rockcastle County', 'Rowan County', 'Shelby County', 'Simpson Cou
nty', 'Spencer County', 'Todd County', 'Trigg County', 'Trimble County', 'Whitley County', 'Wolfe County', 'S
t. John the Baptist Parish', 'St. Landry Parish', 'St. Martin Parish', 'St. Mary Parish', 'St. Tammany Paris
h', 'Tangipahoa Parish', 'Tensas Parish', 'Terrebonne Parish', 'Union Parish', 'Vermilion Parish', 'Vernon Pa
rish', 'Washington Parish', 'Webster Parish', 'West Baton Rouge Parish', 'West Carroll Parish', 'West Felicia
na Parish', 'Winn Parish', 'Androscoggin County', 'Aroostook County', 'Cumberland County', 'Kennebec County',
'Oxford County', 'Penobscot County', 'Piscataquis County', 'Sagadahoc County', 'Somerset County', 'Waldo Coun
ty', 'York County', 'Allegany County', 'Anne Arundel County', 'Baltimore County', 'Calvert County', 'Caroline
County', 'Cecil County', 'Charles County', 'Dorchester County', 'Frederick County', 'Garrett County', 'Harfor
```

d County', 'Gogebic County', 'Grand Traverse County', 'Gratiot County', 'Hillsdale County', 'Houghton County', 'Huron County', 'Ingham County', 'Ionia County', 'Iosco County', 'Iron County', 'Isabella County', 'Kalamazoo County', 'Kalkaska County', 'Keweenaw County', 'Lapeer County', 'Leelanau County', 'Lenawee County', 'Luce County', 'Mackinac County', 'Macomb County', 'Manistee County', 'Marquette County', 'Mecosta County', 'Menominee County', 'Midland County', 'Missaukee County', 'Montcalm County', 'Montmorency County', 'Muskegon County', 'Newaygo County', 'Oakland County', 'Oceana County', 'Ogemaw County', 'Ontonagon County', 'Oscoda County', 'Otsego County', 'Ottawa County', 'Presque Isle County', 'Roscommon County', 'Saginaw County', 'St. Clair County', 'McLeod County', 'Mahnomen County', 'Meeker County', 'Mille Lacs County', 'Morrison County', 'Mower County', 'Murray County', 'Nicollet County', 'Nobles County', 'Norman County', 'Olmsted County', 'Otter Tail County', 'Pennington County', 'Pine County', 'Pipestone County', 'Ramsey County', 'Red Lake County', 'Redwood County', 'Renville County', 'Rock County', 'Roseau County', 'St. Louis County', 'Sherburne County', 'Sibley County', 'Stearns County', 'Steele County', 'Swift County', 'Traverse County', 'Wabasha County', 'Wadena County', 'Waseca County', 'Watonwan County', 'Wilkin County', 'Winona County', 'Wright County', 'Yellow Medicine County', 'Alcorn County', 'Amite County', 'Oktibbeha County', 'Panola County', 'Pearl River County', 'Pontotoc County', 'Prentiss County', 'Quitman County', 'Rankin County', 'Sharkey County', 'Stone County', 'Sunflower County', 'Tallahatchie County', 'Tate County', 'Tippah County', 'Tishomingo County', 'Tunica County', 'Walthall County', 'Winston County', 'Yalobusha County', 'Yazoo County', 'Andrew County', 'Atchison County', 'Audrain County', 'Barry County', 'Barton County', 'Bates County', 'Bollinger County', 'Bucha County', 'Callaway County', 'Cape Girardeau County', 'Carter County', 'Cass County', 'New Madrid County', 'Nodaway County', 'Oregon County', 'Osage County', 'Ozark County', 'Pemiscot County', 'Pettis County', 'Phelps County', 'Platte County', 'Ralls County', 'Ray County', 'Reynolds County', 'St. Charles County', 'Ste. Genevieve County', 'St. Francois County', 'Schuyler County', 'Scotland County', 'Shannon County', 'Stoddard County', 'Sullivan County', 'Taney County', 'Texas County', 'Vernon County', 'St. Louis city', 'Beaverhead County', 'Big Horn County', 'Treasure County', 'Valley County', 'Wheatland County', 'Wibaux County', 'Yellowstone County', 'Antelope County', 'Arthur County', 'Banner County', 'Box Butte County', 'Brown County', 'Buffalo County', 'Burt County', 'Cedar County', 'Chase County', 'Cherry County', 'Cheyenne County', 'Colfax County', 'Cuming County', 'Dakota County', 'Dawes County', 'Deuel County', 'Dixon County', 'Dundy County', 'Fillmore County', 'Frontier County', 'Furnas County', 'Gage County', 'Garden County', 'Garfield County', 'Gosper County', 'Greeley County', 'Harlan County', 'Hayes County', 'Hitchcock County', 'Holt County', 'Hooker County', 'Elko County', 'Esmeralda County', 'Eureka County', 'Lander County', 'Nye County', 'Pershing County', 'Storey County', 'Washoe County', 'White Pine County', 'Carson City', 'Belknap County', 'Cheshire County', 'Coos County', 'Grafton County', 'Hillsborough County', 'Merrimack County', 'Rockingham County', 'Strafford County', 'Atlantic County', 'Bergen County', 'Burlington County', 'Cape May County', 'Essex County', 'Gloucester County', 'Hudson County', 'Hunterdon County', 'Monmouth County', 'Morris County', 'Ocean County', 'Passaic County', 'Salem County', 'Bernalillo County', 'Catron County', 'Chaves County', 'Cibola County', 'Herkimer County', 'Lewis County', 'Nassau County', 'New York County', 'Niagara County', 'Oneida County', 'Onondaga County', 'Ontario County', 'Orleans County', 'Oswego County', 'Queens County', 'Rensselaer County', 'Richmond County', 'Rockland County', 'St. Lawrence County', 'Saratoga County', 'Schenectady County', 'Schoharie County', 'Seneca County', 'Steuben County', 'Suffolk County', 'Tioga County', 'Tompkins County', 'Ulster County', 'Westchester County', 'Wyoming County', 'Yates County', 'Alamance County', 'Alexander County', 'Alleghany County', 'Anson County', 'Avery County', 'Beaufort County', 'Bertie County', 'Mecklenburg County', 'Moore County', 'Nash County', 'New Hanover County', 'Northampton County', 'Onslow County', 'Pamlico County', 'Pasquotank County', 'Pender County', 'Perquimans County', 'Person County', 'Pitt County', 'Robeson County', 'Rutherford County', 'Sampson County', 'Stanly County', 'Stokes County', 'Surry County', 'Swain County', 'Transylvania County', 'Tyrrell County', 'Vance County', 'Wake County', 'Watauga County', 'Yadkin County', 'Yancey County', 'Barnes County', 'Benson County', 'Billings County', 'Bottineau County', 'Bowman County', 'Ashtabula County', 'Athens County', 'Auglaize County', 'Belmont County', 'Champaign County', 'Clermont County', 'Clinton County', 'Columbiana County', 'Coshocton County', 'Cuyahoga County', 'Darke County', 'Defiance County', 'Delaware County', 'Erie County', 'Gallia County', 'Geauga County', 'Guernsey County', 'Highland County', 'Hocking County', 'Holmes County', 'Licking County', 'Lorain County', 'Mahoning County', 'Medina County', 'Coal County', 'Comanche County', 'Cotton County', 'Craig County', 'Creek County', 'Dewey County', 'Ellis County', 'Garvin County', 'Greer County', 'Harmon County', 'Harper County', 'Haskell County', 'Hughes County', 'Johnston County', 'Kay County', 'Kingfisher County', 'Kiowa County', 'Latimer County', 'Le Flore County', 'Love County', 'McClain County', 'McCurtain County', 'McIntosh County', 'Major County', 'Mayes County', 'Muskogee County', 'Nowata County', 'Okfuskee County', 'Oklahoma County', 'Okmulgee County', 'Pawnee County', 'Payne County', 'Pittsburg County', 'Yamhill County', 'Allegheny County', 'Armstrong County', 'Beaver County', 'Bedford County', 'Berks County', 'Blair County', 'Bucks County', 'Cambria County', 'Cameron County', 'Carbon County', 'Centre County', 'Chester County', 'Clarion County', 'Clearfield County', 'Dauphin County', 'Elk County', 'Forest County', 'Huntingdon County', 'Indiana County', 'Juniata County', 'Lackawanna County', 'Lancaster County', 'Lebanon County', 'Lehigh County', 'Luzerne County', 'Lycoming County', 'McKean County', 'Mifflin County', 'Montour County', 'Greenwood County', 'Hampton County', 'Horry County', 'Kershaw County', 'Lexington County', 'McCormick County', 'Marlboro County', 'Newberry County', 'Oconee County', 'Orangeburg County', 'Pickens County', 'Richland County', 'Saluda County', 'Spartanburg County', 'Williamsburg County', 'Aurora County', 'Beadle County', 'Bennett County', 'Bon Homme County', 'Brookings County', 'Brule County', 'Charles Mix County', 'Codington County', 'Corson County', 'Davison County', 'Day County', 'Edmunds County', 'Fall River County', 'Faulk County', 'Gregory County', 'Cheatham County', 'Claiborne County', 'Cocke County', 'Crockett County', 'Davidson County', 'Dickson County', 'Dyer County', 'Fentress County', 'Giles County', 'Grainger County', 'Hamblen County', 'Hardeman County', 'Hawkins County', 'Haywood County', 'Hickman County', 'Humphreys County', 'Lauderdale County', 'Loudon County', 'McMinn County', 'McNairy County', 'Bee County', 'Bexar County', 'Blanco County', 'Borden County', 'Bosque County', 'Bowie County', 'Brazoria County', 'Brazos County', 'Brewster County', 'Briscoe County', 'Burleson County', 'Burnet County', 'Callahan County', 'Camp County', 'Carson County', 'Castro County', 'Childress County', 'Cochran County', 'Coke County', 'Coleman County', 'Collin County', 'Collingsworth County', 'Colorado County', 'Comal County', 'Concho County', 'Cooke County', 'Coryell County', 'Cottle County', 'Crane County', 'Crosby County', 'Culberson County', 'Dallam County', 'Deaf Smith County', 'Delta County', 'Denton County', 'Hood County', 'Hopkins County', 'Hudspeth County', 'Hunt County', 'Hutchinson County', 'Irion County', 'Jack County', 'Jim Hogg County', 'Jim Wells County', 'Karnes County', 'Kaufman County', 'Kenedy County', 'Kerr County', 'Kimble County', 'King County', 'Kinney County', 'Kleberg County', 'Lamb County', 'Lampasas County', 'La Salle County', 'Lavaca County', 'Leon County', 'Liberty County', 'Limestone County', 'Lipscomb County', 'Live Oak County', 'Llano County', 'Loving County', 'Lubbock County', 'Lynn County', 'McCulloch County', 'McLen County', 'McMullen County', 'Somervell County', 'Starr County', 'Stephens County', 'Sterling County', 'Stonewall County', 'Sutton County', 'Swisher County', 'Tarrant County', 'Terrell County', 'Terry County', 'Throckmorton County', 'Titus County', 'Tom Green County', 'Travis County', 'Trinity County', 'Tyler County', 'Upshur County', 'Upton

County', 'Uvalde County', 'Val Verde County', 'Van Zandt County', 'Victoria County', 'Waller County', 'Ward County', 'Webb County', 'Wharton County', 'Wilbarger County', 'Willacy County', 'Williamson County', 'Winkler County', 'Wise County', 'Wood County', 'Yoakum County', 'Young County', 'Zapata County', 'Zavala County', 'Box Elder County', 'Cache County', 'Bland County', 'Botetourt County', 'Brunswick County', 'Buckingham County', 'Charles City County', 'Charlotte County', 'Chesterfield County', 'Culpeper County', 'Dickenson County', 'Dinwiddie County', 'Fairfax County', 'Fauquier County', 'Fluvanna County', 'Goochland County', 'Grayson County', 'Greensville County', 'Halifax County', 'Hanover County', 'Henrico County', 'Isle of Wight County', 'James City County', 'King and Queen County', 'King George County', 'King William County', 'Loudoun County', 'Lunenburg County', 'Mathews County', 'Franklin city', 'Fredericksburg city', 'Galax city', 'Hampton city', 'Harrisonburg city', 'Hopewell city', 'Lexington city', 'Lynchburg city', 'Manassas city', 'Manassas Park city', 'Martinsville city', 'Newport News city', 'Norfolk city', 'Norton city', 'Petersburg city', 'Poquoson city', 'Portsmouth city', 'Radford city', 'Richmond city', 'Roanoke city', 'Salem city', 'Staunton city', 'Suffolk city', 'Virginia Beach city', 'Waynesboro city', 'Williamsburg city', 'Winchester city', 'Asotin County', 'Chelan County', 'Clallam County', 'Cowlitz County', 'Ferry County', 'Grays Harbor County', 'Island County', 'Kitsap County', 'Kittitas County', 'Klickitat County', 'Pleasants County', 'Preston County', 'Raleigh County', 'Ritchie County', 'Roane County', 'Summers County', 'Tucker County', 'Wetzel County', 'Wirt County', 'Ashland County', 'Barron County', 'Bayfield County', 'Burnett County', 'Calumet County', 'Chippewa County', 'Dane County', 'Door County', 'Dunn County', 'Eau Claire County', 'Florence County', 'Fond du Lac County', 'Green County', 'Green Lake County', 'Converse County', 'Crook County', 'Goshen County', 'Hot Springs County', 'Laramie County', 'Natrona County', 'Niobrara County', 'Sublette County', 'Sweetwater County', 'Teton County', 'Uinta County', 'Washakie County', 'Weston County', 'Lowndes County', 'Marengo County', 'Mobile County', 'Talladega County', 'Tallapoosa County', 'Tuscaloosa County', 'Aleutians East Borough', 'Aleutians West Census Area', 'Anchorage Municipality', 'Bethel Census Area', 'Bristol Bay Borough', 'Denali Borough', 'Dillingham Census Area', 'Fairbanks North Star Borough', 'Haines Borough', 'Hoonah-Angoon Census Area', 'Juneau City and Borough', 'Kenai Peninsula Borough', 'Ketchikan Gateway Borough', 'Kodiak Island Borough', 'Kusilvak Census Area', 'Drew County', 'Faulkner County', 'Garland County', 'Hempstead County', 'Hot Spring County', 'Independence County', 'Izard County', 'Lafayette County', 'Little River County', 'Lonoke County', 'Miller County', 'Mississippi County', 'Ouachita County', 'Poinsett County', 'Prairie County', 'St. Francis County', 'Searcy County', 'Sebastian County', 'Sevier County', 'Sharp County', 'Shasta County', 'Sierra County', 'Siskiyou County', 'Solano County', 'Sonoma County', 'Stanislaus County', 'Sutter County', 'Tehama County', 'Tulare County', 'Tuolumne County', 'Ventura County', 'Yolo County', 'Yuba County', 'Alamosa County', 'Arapahoe County', 'Archuleta County', 'Baca County', 'Bent County', 'Boulder County', 'Broomfield County', 'Chaffee County', 'Clear Creek County', 'Conejos County', 'Costilla County', 'Crowley County', 'Denver County', 'Dolores County', 'Eagle County', 'El Paso County', 'Gilpin County', 'Grand County', 'Gunnison County', 'Hinsdale County', 'Huerfano County', 'Kit Carson County', 'Citrus County', 'Collier County', 'DeSoto County', 'Dixie County', 'Duval County', 'Flagler County', 'Gadsden County', 'Gilchrist County', 'Glades County', 'Gulf County', 'Hardee County', 'Hendry County', 'Herdo County', 'Highlands County', 'Indian River County', 'Levy County', 'Manatee County', 'Miami-Dade County', 'Okaloosa County', 'Okeechobee County', 'Palm Beach County', 'Pasco County', 'Pinellas County', 'St. Johns County', 'St. Lucie County', 'Long County', 'Lumpkin County', 'McDuffie County', 'Meriwether County', 'Muscogee County', 'Oglethorpe County', 'Paulding County', 'Peach County', 'Pierce County', 'Rabun County', 'Rockdale County', 'Schley County', 'Screven County', 'Spalding County', 'Stewart County', 'Talbot County', 'Taliaferro County', 'Tattnall County', 'Telfair County', 'Gem County', 'Gooding County', 'Idaho County', 'Jerome County', 'Kootenai County', 'Latah County', 'Lemhi County', 'Minidoka County', 'Nez Perce County', 'Owyhee County', 'Payette County', 'Power County', 'Shoshone County', 'Twin Falls County', 'Bond County', 'Bureau County', 'Christian County', 'Coles County', 'De Witt County', 'DuPage County', 'Edgar County', 'Edwards County', 'Ford County', 'Rock Island County', 'Sangamon County', 'Stark County', 'Stephenson County', 'Tazewell County', 'Vermilion County', 'Wabash County', 'Whiteside County', 'Will County', 'Winnebago County', 'Woodford County', 'Bartholomew County', 'Blackford County', 'Daviess County', 'Dearborn County', 'Dubois County', 'Elkhart County', 'Fountain County', 'Starke County', 'Switzerland County', 'Tippecanoe County', 'Tipton County', 'Vanderburgh County', 'Vermillion County', 'Vigo County', 'Warrick County', 'Wells County', 'Allamakee County', 'Appanoose County', 'Audubon County', 'Black Hawk County', 'Bremer County', 'Buena Vista County', 'Cerro Gordo County', 'Chickasaw County', 'Davis County', 'Des Moines County', 'Poweshiek County', 'Ringgold County', 'Sac County', 'Sioux County', 'Story County', 'Tama County', 'Wapello County', 'Winneshiek County', 'Woodbury County', 'Barber County', 'Chautauqua County', 'Cloud County', 'Coffey County', 'Cowley County', 'Doniphan County', 'Ellsworth County', 'Finney County', 'Geary County', 'Gove County', 'Gray County', 'Harvey County', 'Hodgeman County', 'Jewell County', 'Kearny County', 'Kingman County', 'Labette County', 'Lane County', 'Leavenworth County', 'McPherson County', 'Morton County', 'Nemaha County', 'Neosho County', 'Ness County', 'Norton County', 'Osborne County', 'Pottawatomie County', 'Pratt County', 'Rawlins County', 'Casey County', 'Edmonson County', 'Elliott County', 'Estill County', 'Fleming County', 'Garrard County', 'Graves County', 'Greenup County', 'Jessamine County', 'Kenton County', 'Knott County', 'Larue County', 'Laurel County', 'Leslie County', 'Letcher County', 'Acadia Parish', 'Allen Parish', 'Ascension Parish', 'Assumption Parish', 'Avoyelles Parish', 'Beauregard Parish', 'Bienville Parish', 'Bossier Parish', 'Caddo Parish', 'Calcasieu Parish', 'Caldwell Parish', 'Cameron Parish', 'Catahoula Parish', 'Claiborne Parish', 'Concordia Parish', 'De Soto Parish', 'East Baton Rouge Parish', 'East Carroll Parish', 'East Feliciana Parish', 'Evangeline Parish', 'Franklin Parish', 'Grant Parish', 'Iberia Parish', 'Iberville Parish', 'Jackson Parish', 'Jefferson Parish', 'Jefferson Davis Parish', 'Lafayette Parish', 'Lafourche Parish', 'LaSalle Parish', 'Lincoln Parish', 'Livingston Parish', 'Madison Parish', 'Morehouse Parish', 'Natchitoches Parish', 'Orleans Parish', 'Ouachita Parish', 'Plaquemines Parish', 'Pointe Coupee Parish', 'Rapides Parish', 'Red River Parish', 'Richland Parish', 'Sabine Parish', 'St. Bernard Parish', 'St. Charles Parish', 'St. Helena Parish', 'St. James Parish', "Prince George's County", "Queen Anne's County", "St. Mary's County", 'Wicomico County', 'Worcester County', 'Baltimore city', 'Barnstable County', 'Berkshire County', 'Bristol County', 'Dukes County', 'Hampden County', 'Hampshire County', 'Nantucket County', 'Norfolk County', 'Alcona County', 'Alger County', 'Allegan County', 'Alpena County', 'Antrim County', 'Arenac County', 'Baraga County', 'Benzie County', 'Branch County', 'Charlevoix County', 'Cheboygan County', 'Clare County', 'Eaton County', 'Genesee County', 'Gladwin County', 'Sanilac County', 'Schoolcraft County', 'Shiawassee County', 'Tuscola County', 'Washtenaw County', 'Wexford County', 'Aitkin County', 'Anoka County', 'Becker County', 'Beltrami County', 'Big Stone County', 'Blue Earth County', 'Carlton County', 'Carver County', 'Chisago County', 'Cottonwood County', 'Crow Wing County', 'Faribault County', 'Freeborn County', 'Goodhue County', 'Hennepin County', 'Hubbard County', 'Isanti County', 'Itasca County', 'Kanabec County', 'Kandiyohi County', 'Kittson County', 'Koochiching County', 'Lac qui Parle County', 'Lake of the Woods County', 'Le Sueur County', 'Attala County', 'Bolivar County', 'Coahoma County', 'Copiah County', 'Forrest County', 'George County', 'Grenada County', 'Hinds County', 'Issaquena County', 'Itawamba County', 'Jefferson Davis County', 'Kemper County', 'Leake County', 'Leflore County', 'Neshoba County', 'Noxubee County', 'Charito

n County', 'Cole County', 'Cooper County', 'Dent County', 'Dunklin County', 'Gasconade County', 'Gentry County', 'Hickory County', 'Howell County', 'Laclede County', 'McDonald County', 'Maries County', 'Moniteau County', 'Broadwater County', 'Cascade County', 'Chouteau County', 'Daniels County', 'Deer Lodge County', 'Fallon County', 'Fergus County', 'Flathead County', 'Glacier County', 'Golden Valley County', 'Granite County', 'Hill County', 'Judith Basin County', 'Lewis and Clark County', 'McCone County', 'Meagher County', 'Missoula County', 'Musselshell County', 'Petroleum County', 'Pondera County', 'Powder River County', 'Ravalli County', 'Roosevelt County', 'Rosebud County', 'Sanders County', 'Silver Bow County', 'Stillwater County', 'Sweet Grass County', 'Toole County', 'Kearney County', 'Keith County', 'Keya Paha County', 'Kimball County', 'Loup County', 'Merrick County', 'Morrill County', 'Nance County', 'Nuckolls County', 'Otoe County', 'Perkins County', 'Red Willow County', 'Richardson County', 'Sarpy County', 'Saunders County', 'Scotts Bluff County', 'Thayer County', 'Thurston County', 'Churchill County', 'Curry County', 'De Baca County', 'Eddy County', 'Guadalupe County', 'Harding County', 'Hidalgo County', 'Lea County', 'Los Alamos County', 'Luna County', 'McKinley County', 'Mora County', 'Quay County', 'Rio Arriba County', 'Sandoval County', 'Santa Fe County', 'Socorro County', 'Taos County', 'Torrance County', 'Valencia County', 'Albany County', 'Bronx County', 'Broome County', 'Cattaraugus County', 'Cayuga County', 'Chemung County', 'Chego County', 'Cortland County', 'Dutchess County', 'Bladen County', 'Buncombe County', 'Cabarrus County', 'Carteret County', 'Caswell County', 'Catawba County', 'Chowan County', 'Columbus County', 'Craven County', 'Currituck County', 'Dare County', 'Davie County', 'Duplin County', 'Durham County', 'Edgecombe County', 'Gaston County', 'Gates County', 'Granville County', 'Guilford County', 'Harnett County', 'Hertford County', 'Hoke County', 'Hyde County', 'Iredell County', 'Lenoir County', 'McDowell County', 'Burleigh County', 'Cavalier County', 'Dickey County', 'Divide County', 'Emmons County', 'Foster County', 'Grand Forks County', 'Griggs County', 'Hettinger County', 'Kidder County', 'LaMoure County', 'McKenzie County', 'Mountrail County', 'Oliver County', 'Pembina County', 'Ransom County', 'Rolette County', 'Sargent County', 'Slope County', 'Stutsman County', 'Towner County', 'Traill County', 'Walsh County', 'Williams County', 'Meigs County', 'Morrow County', 'Muskingum County', 'Pickaway County', 'Portage County', 'Preble County', 'Ross County', 'Sandusky County', 'Scioto County', 'Trumbull County', 'Tuscarawas County', 'Van Wert County', 'Vinton County', 'Wyandot County', 'Alfalfa County', 'Atoka County', 'Beckham County', 'Caddo County', 'Canadian County', 'Cimarron County', 'Pushmataha County', 'Roger Mills County', 'Rogers County', 'Sequoyah County', 'Tillman County', 'Tulsa County', 'Wagoner County', 'Washita County', 'Woods County', 'Woodward County', 'Clackamas County', 'Clatsop County', 'Deschutes County', 'Gilliam County', 'Harney County', 'Hood River County', 'Josephine County', 'Klamath County', 'Malheur County', 'Multnomah County', 'Tillamook County', 'Umatilla County', 'Wallowa County', 'Wasco County', 'Northumberland County', 'Philadelphia County', 'Potter County', 'Schuylkill County', 'Snyder County', 'Susquehanna County', 'Vego County', 'Westmoreland County', 'Newport County', 'Providence County', 'Abbeville County', 'Aiken County', 'Allendale County', 'Bamberg County', 'Barnwell County', 'Berkeley County', 'Charleston County', 'Clarendon County', 'Colleton County', 'Darlington County', 'Dillon County', 'Edgefield County', 'Georgetown County', 'Greenville County', 'Haakon County', 'Hamlin County', 'Hand County', 'Hanson County', 'Jerauld County', 'Kingsbury County', 'Lyman County', 'McCook County', 'Mellette County', 'Miner County', 'Minnehaha County', 'Moody County', 'Oglala Lakota County', 'Roberts County', 'Sanborn County', 'Spink County', 'Stanley County', 'Sully County', 'Tripp County', 'Walworth County', 'Yankton County', 'Ziebach County', 'Bledsoe County', 'Cannon County', 'Maury County', 'Obion County', 'Overton County', 'Pickett County', 'Rhea County', 'Sequatchie County', 'Trousdale County', 'Unicoi County', 'Weakley County', 'Andrews County', 'Angelina County', 'Aransas County', 'Archer County', 'Atascosa County', 'Austin County', 'Bailey County', 'Bandera County', 'Bastrop County', 'Baylor County', 'DeWitt County', 'Dickens County', 'Dimmit County', 'Donley County', 'Eastland County', 'Ector County', 'Erath County', 'Falls County', 'Fisher County', 'Foard County', 'Fort Bend County', 'Freestone County', 'Frio County', 'Gaines County', 'Galveston County', 'Garza County', 'Gillespie County', 'Glasscock County', 'Goliad County', 'Gonzales County', 'Gregg County', 'Grimes County', 'Hansford County', 'Hartley County', 'Hays County', 'Hemphill County', 'Hockley County', 'Matagorda County', 'Maverick County', 'Milam County', 'Montague County', 'Motley County', 'Nacogdoches County', 'Navarro County', 'Nolan County', 'Nueces County', 'Ochiltree County', 'Palo Pinto County', 'Parker County', 'Parmer County', 'Pecos County', 'Presidio County', 'Rains County', 'Randall County', 'Reagan County', 'Real County', 'Red River County', 'Reeves County', 'Refugio County', 'Rockwall County', 'Runnels County', 'Rusk County', 'Sabine County', 'San Augustine County', 'San Jacinto County', 'San Patricio County', 'San Saba County', 'Schleicher County', 'Scurry County', 'Shackelford County', 'Daggett County', 'Duchesne County', 'Emery County', 'Juab County', 'Millard County', 'Piute County', 'Rich County', 'Salt Lake County', 'Sanpete County', 'Tooele County', 'Uintah County', 'Utah County', 'Wasatch County', 'Weber County', 'Addison County', 'Bennington County', 'Caledonia County', 'Chittenden County', 'Grand Isle County', 'Lamoille County', 'Rutland County', 'Windsor County', 'Accomack County', 'Albemarle County', 'Amelia County', 'Amherst County', 'Appomattox County', 'Arlington County', 'Augusta County', 'New Kent County', 'Nottoway County', 'Patrick County', 'Pittsylvania County', 'Powhatan County', 'Prince Edward County', 'Prince George County', 'Prince William County', 'Rappahannock County', 'Roanoke County', 'Rockbridge County', 'Shedoah County', 'Smyth County', 'Southampton County', 'Spotsylvania County', 'Wythe County', 'Alexandria city', 'Bristol city', 'Buena Vista city', 'Charlottesville city', 'Chesapeake city', 'Colonial Heights city', 'Covington city', 'Danville city', 'Emporia city', 'Fairfax city', 'Falls Church city', 'Okanogan County', 'Pacific County', 'Pend Oreille County', 'Skagit County', 'Skamania County', 'Snohomish County', 'Spokane County', 'Wahkiakum County', 'Walla Walla County', 'Whatcom County', 'Whitman County', 'Yakima County', 'Braxton County', 'Brooke County', 'Cabell County', 'Doddridge County', 'Greenbrier County', 'Hardy County', 'Kanawha County', 'Mingo County', 'Monongalia County', 'Juneau County', 'Kenosha County', 'Kewaunee County', 'La Crosse County', 'Langlade County', 'Manitowoc County', 'Marathon County', 'Marinette County', 'Milwaukee County', 'Oconto County', 'Outagamie County', 'Ozaukee County', 'Pepin County', 'Price County', 'Racine County', 'St. Croix County', 'Sauk County', 'Sawyer County', 'Shawano County', 'Sheboygan County', 'Trempealeau County', 'Vilas County', 'Washburn County', 'Waukesha County', 'Waupaca County', 'Waushara County', 'Doña Ana County', nan]

Observations so far:

- There are 1,877 different counties, with Washington County being the most common.
- It looks like we have an outlier of -38 for "Population Not Staying at Home" and -108 for "Number of Trips 3-5" which doesn't make sense. Negative numbers will need to be removed.
- There may be some missing data (NaN)

Missing Data

```
In [8]:  # Checking missing data sums
         trip_data.isna().sum()
```

```
Out[8]:  Level                             0
         Date                              0
         State FIPS                      663
         State Postal Code               663
         County FIPS                   34476
         County Name                   34476
         Population Staying at Home    22945
         Population Not Staying at Home 22945
         Number of Trips              22945
         Number of Trips <1           22945
         Number of Trips 1-3          22945
         Number of Trips 3-5          22945
         Number of Trips 5-10         22945
         Number of Trips 10-25        22945
         Number of Trips 25-50        22945
         Number of Trips 50-100       22945
         Number of Trips 100-250      22945
         Number of Trips 250-500      22945
         Number of Trips >=500        22945
         dtype: int64
```

Since we have a good amount of data to work with, we're going to remove the rows with missing data.

Data Cleanup

```
In [9]:  # Removing missing data
         trip_data_clean = trip_data.dropna()

         # Checking missing data sums
         trip_data_clean.isna().sum()
```

```
Out[9]:  Level                           0
         Date                            0
         State FIPS                      0
         State Postal Code              0
         County FIPS                     0
         County Name                     0
         Population Staying at Home      0
         Population Not Staying at Home  0
         Number of Trips                 0
         Number of Trips <1              0
         Number of Trips 1-3             0
         Number of Trips 3-5             0
         Number of Trips 5-10            0
         Number of Trips 10-25           0
         Number of Trips 25-50           0
         Number of Trips 50-100          0
         Number of Trips 100-250         0
         Number of Trips 250-500         0
         Number of Trips >=500           0
         dtype: int64
```

```
In [10]: # Remove outliers (negative numbers)

         trip_data_clean = trip_data_clean[trip_data_clean['Population Not Staying at Home'] > -1]
```

```
In [11]:  # Converting dates to datetime format
          trip_data_clean['Date'] = pd.to_datetime(trip_data_clean['Date'])

          trip_data_clean['Date'].head(10)
```

Out[11]: 0   2019-01-01
         2   2019-01-01
         3   2019-01-01
         4   2019-01-01
         5   2019-01-01
         6   2019-01-01
         7   2019-01-01
         8   2019-01-01
         9   2019-01-01
         10   2019-01-01
         Name: Date, dtype: datetime64[ns]

```
In [12]:  # Summary information for datetime column
          print("Datetime Data:\n")
          print(trip_data_clean['Date'].describe())

          # Summary information for cleaned string columns
          print("String Data:\n")
          print(trip_data_clean.describe(include=['O']))

          # Summary information for cleaned numeric columns without scientific notation
          with pd.option_context('float_format', '{:f}'.format):
              print("\nNumeric Data:\n\n", trip_data_clean.describe())
```

```
Datetime Data:

count                  2060200
unique                     663
top        2020-07-22 00:00:00
freq                      3141
first      2019-01-01 00:00:00
last       2020-10-24 00:00:00
Name: Date, dtype: object
String Data:

        Level State Postal Code        County Name
count  2060200       2060200           2060200
unique       1            51              1876
top     County            TX   Washington County
freq   2060200        165599             19890

Numeric Data:

           State FIPS     County FIPS  Population Staying at Home  \
count  2060200.000000  2060200.000000             2060200.000000
mean        30.291433    30395.011147                22564.431223
std         15.142546    15160.851846                77587.660701
min          1.000000     1001.000000                    8.000000
25%         18.000000    18177.000000                 2101.000000
50%         29.000000    29163.000000                 4889.000000
75%         45.000000    45081.000000                13758.000000
max         56.000000    56045.000000              3644862.000000

        Population Not Staying at Home  Number of Trips  Number of Trips <1  \
count                  2060200.000000   2060200.000000      2060200.000000
mean                     82528.591029    386462.136244        94520.003898
std                     259854.539076   1227592.514024       330072.548121
min                        114.000000       220.000000            0.000000
25%                       9148.000000     42030.000000         8952.000000
50%                      21307.000000     99985.000000        21547.000000
75%                      55802.000000    264720.000000        58324.000000
max                    8636354.000000  54507586.000000     14779760.000000

        Number of Trips 1-3  Number of Trips 3-5  Number of Trips 5-10  \
count        2060200.000000       2060200.000000        2060200.000000
mean           97004.714446         47359.270296          59889.925995
std           307381.899139        149973.967813         193136.899844
min                0.000000             0.000000              0.000000
25%             9449.000000          3974.000000           5464.000000
50%            24759.000000         11339.000000          13992.000000
75%            68442.000000         32827.000000          39102.000000
max         16044708.000000       7175390.000000        8268313.000000

        Number of Trips 10-25  Number of Trips 25-50  Number of Trips 50-100  \
count         2060200.000000         2060200.000000          2060200.000000
mean            59090.957052           19019.368381             6020.042833
std            186060.325657           52003.898772            14465.956313
min                 0.000000               0.000000                0.000000
25%              6683.000000            3309.000000             1149.000000
50%             15836.000000            7074.000000             2509.000000
75%             40460.000000           16272.000000             5646.000000
max           7590739.000000         2117534.000000           610909.000000

        Number of Trips 100-250  Number of Trips 250-500  Number of Trips >=500
count           2060200.000000           2060200.000000         2060200.000000
mean               2465.361245               569.128064             523.364035
std                6259.516351              1696.228660            2211.904276
min                   0.000000                 0.000000               0.000000
25%                 364.000000                55.000000              18.000000
50%                 850.000000               162.000000              66.000000
75%                2101.000000               444.000000             240.000000
max              321128.000000             83066.000000          204712.000000
```
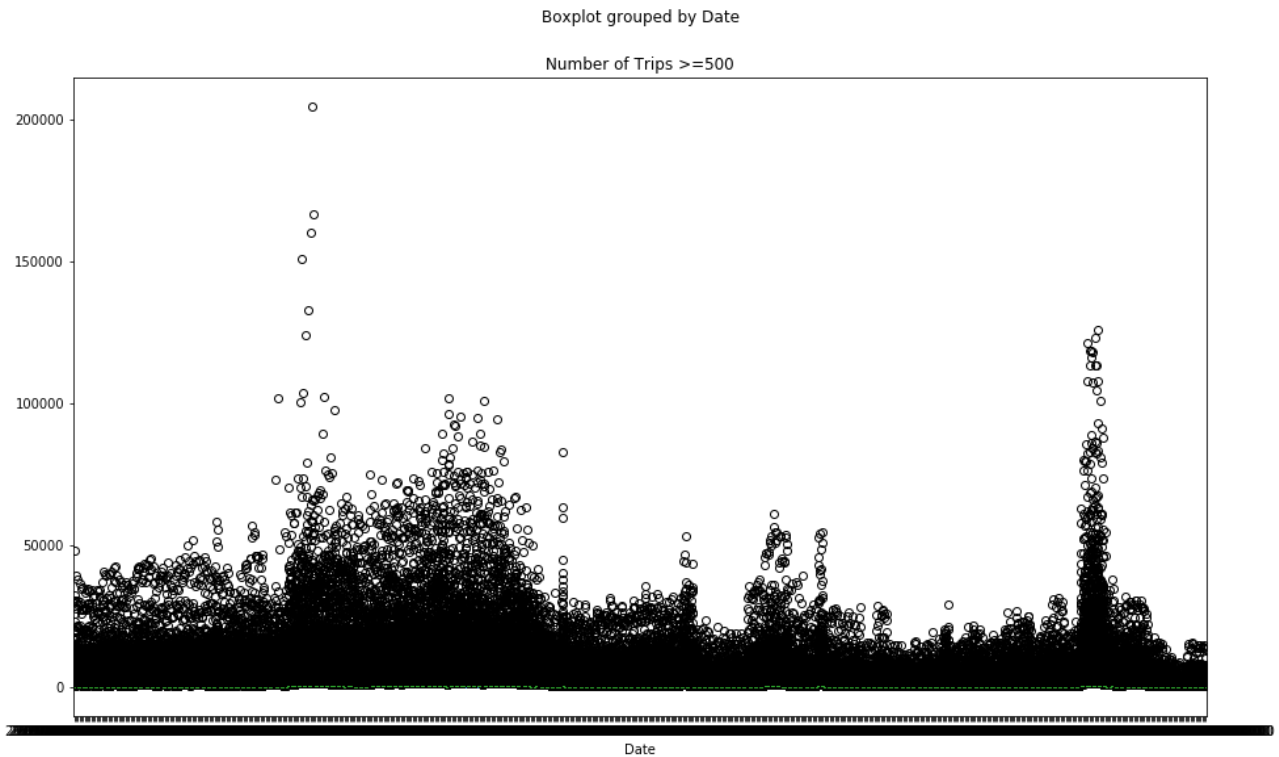
With the newly updated data, we now have observations up until October 24, 2020. The previous version of the data set only went up until August 29th.

Exploratory Visualizations: All Data

```
In [13]:  # Boxplot of number of trips above 500 miles
          trip_data_clean.boxplot(by ='Date',
                                  column =['Number of Trips >=500'],
                                  figsize=(15,9),
                                  grid = False)
```

Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x1182957f0>

Boxplot grouped by Date

Number of Trips >=500

Using all of the observations in one graph is mostly convoluted and ineffective, so in order to make our data more managable and focused, we're going to narrow our analysis to California counties. Even amidst the overcrowding in the above plot, we can see the general ups and downs of travel displayed. The trend is what we originally expected. There are more trips in the first half of the graph (2019) with a drop across most of 2020, and a dramatic spike during the summer (August 2020). Finally, there is a drop at the end of the data during the Fall. To better explain this trend, we might need to supplement with Covid-19 data.

Filtering Data to California Counties

```
In [14]:  # Filtering to California counties
          trip_data_clean_CA = trip_data_clean[trip_data_clean['State Postal Code']=='CA']

          # Previewing filtered data
          print('Shape:', trip_data_clean_CA.shape)
          trip_data_clean_CA.head(5)
```

Shape: (38093, 19)

Out[14]:

	Level	Date	State FIPS	State Postal Code	County FIPS	County Name	Population Staying at Home	Population Not Staying at Home	Number of Trips	Number of Trips <1	Number of Trips 1-3	Number of Trips 3-5	Number of Trips 5-10	Number of Trips 10-25	Number of Trips 25-5
89	County	2019-01-01	6.0	CA	6001.0	Alameda County	387930.0	1273729.0	4689018.0	1498736.0	1167294.0	511906.0	547995.0	659499.0	216985.
90	County	2019-01-01	6.0	CA	6003.0	Alpine County	328.0	770.0	2692.0	629.0	686.0	149.0	196.0	218.0	506.
91	County	2019-01-01	6.0	CA	6005.0	Amador County	9453.0	29810.0	98532.0	13691.0	23586.0	14228.0	12193.0	18232.0	10629.
92	County	2019-01-01	6.0	CA	6007.0	Butte County	57517.0	173032.0	642628.0	178158.0	203203.0	87995.0	64251.0	55810.0	32636.
93	County	2019-01-01	6.0	CA	6009.0	Calaveras County	9464.0	35999.0	107948.0	14325.0	19901.0	14581.0	19746.0	22491.0	10424.

Previously, there were 34,863 observations for California, so with the updated data we've gained 3,230 more observations.

```
In [14]:  # Filtering to California counties
          trip_data_clean_CA = trip_data_clean[trip_data_clean['State Postal Code']=='CA']

          # Previewing filtered data
          print('Shape:', trip_data_clean_CA.shape)
```

```
In [15]:  # Summary information for datetime column
          print("Datetime Data:\n")
          print(trip_data_clean_CA['Date'].describe())

          # Summary information for cleaned string columns
          print("\nString Data:\n")
          print(trip_data_clean_CA.describe(include=['O']))

          # Summary information for cleaned numeric columns without scientific notation
          with pd.option_context('float_format', '{:f}'.format):
              print("\nNumeric Data:\n\n", trip_data_clean_CA.describe())
```

Datetime Data:

```
count                   38093
unique                    663
top       2019-02-04 00:00:00
freq                       58
first     2019-01-01 00:00:00
last      2020-10-24 00:00:00
Name: Date, dtype: object
```

String Data:

```
       Level State Postal Code   County Name
count  38093            38093         38093
unique     1                1            58
top    County               CA  Lassen County
freq   38093            38093           663
```

Numeric Data:

```
        State FIPS  County FIPS  Population Staying at Home  \
count 38093.000000 38093.000000               38093.000000
mean      6.000000  6058.269420              154315.019951
std       0.000000    33.293428              335062.554489
min       6.000000  6001.000000                 153.000000
25%       6.000000  6029.000000               10581.000000
50%       6.000000  6059.000000               44414.000000
75%       6.000000  6087.000000              150289.000000
max       6.000000  6115.000000             3644862.000000

       Population Not Staying at Home  Number of Trips  Number of Trips <1  \
count                    38093.000000     38093.000000        38093.000000
mean                    532992.032473   2416024.906650       630633.083506
std                    1141671.499593   5290218.420634      1391725.358638
min                        708.000000      2199.000000           72.000000
25%                      37795.000000    160152.000000        34269.000000
50%                     148528.000000    686519.000000       176061.000000
75%                     559732.000000   2272921.000000       602246.000000
max                    8636354.000000  54507586.000000     14779760.000000

       Number of Trips 1-3  Number of Trips 3-5  Number of Trips 5-10  \
count         38093.000000         38093.000000          38093.000000
mean         623724.072533        281803.323865         349058.419893
std         1345338.066831        639695.428297         820250.919018
min               0.000000             0.000000              0.000000
25%           41680.000000         17327.000000          21079.000000
50%          192541.000000         77476.000000          75651.000000
75%          617023.000000        266492.000000         303783.000000
max        16044708.000000       7175390.000000        8268313.000000

       Number of Trips 10-25  Number of Trips 25-50  Number of Trips 50-100  \
count           38093.000000           38093.000000            38093.000000
mean           348411.820098          122845.960990            40431.976006
std            792737.112743          238426.804016            70671.404309
min                 0.000000               0.000000                0.000000
25%             24120.000000           11828.000000             4880.000000
50%             85969.000000           41361.000000            13093.000000
75%            284167.000000          116216.000000            44619.000000
max           7590739.000000         2117534.000000           610909.000000

       Number of Trips 100-250  Number of Trips 250-500  Number of Trips >=500
count             38093.000000             38093.000000           38093.000000
mean              13145.414643              2970.886436            2999.948678
std               24104.216492              6122.605376            6873.082569
min                   0.000000                 0.000000               0.000000
25%                1366.000000               229.000000             153.000000
50%                5203.000000               959.000000             720.000000
75%               12816.000000              2654.000000            2572.000000
max              321128.000000             83066.000000          125691.000000
```

```
# List unique values in the 'County Name' column
counties_list_CA = trip_data_clean_CA['County Name'].unique().tolist()

print(counties_list_CA)
```
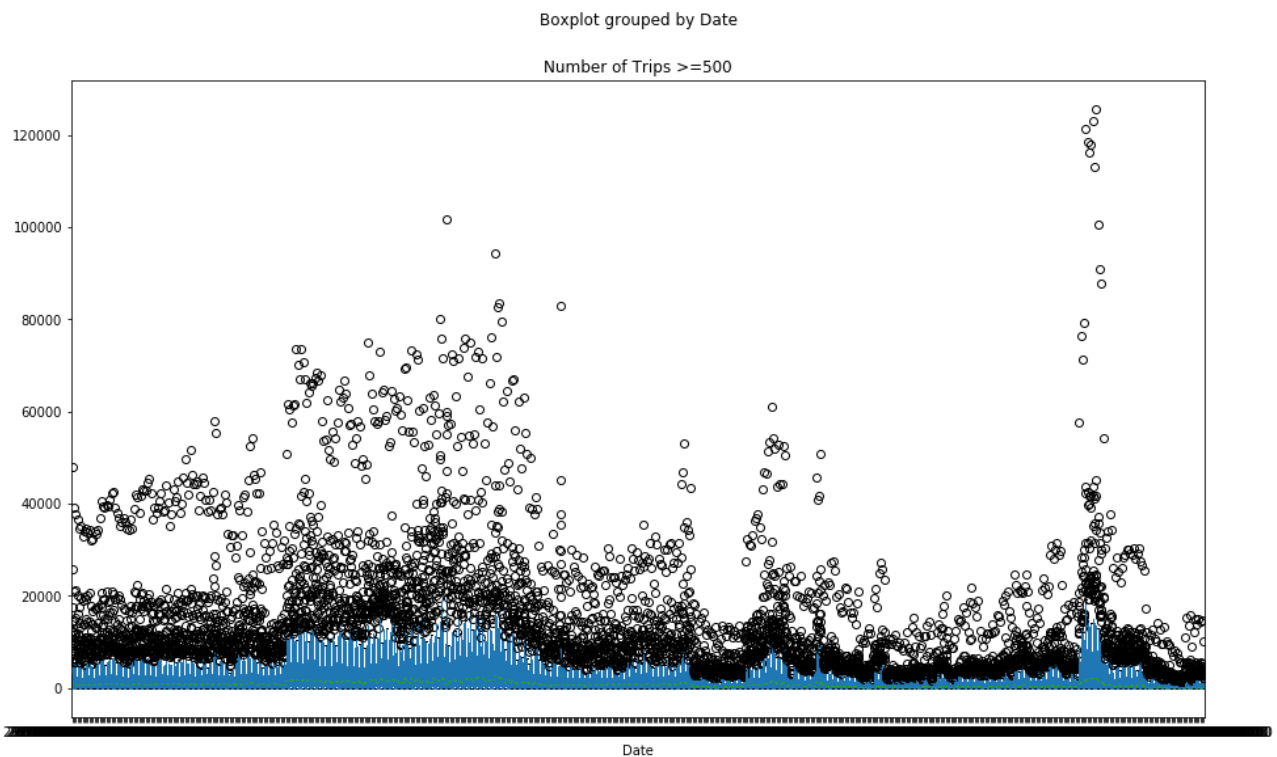
['Alameda County', 'Alpine County', 'Amador County', 'Butte County', 'Calaveras County', 'Colusa County', 'Co
ntra Costa County', 'Del Norte County', 'El Dorado County', 'Fresno County', 'Glenn County', 'Humboldt Count
y', 'Imperial County', 'Inyo County', 'Kern County', 'Kings County', 'Lake County', 'Lassen County', 'Los Ang
eles County', 'Madera County', 'Marin County', 'Mariposa County', 'Mendocino County', 'Merced County', 'Modoc
County', 'Mono County', 'Monterey County', 'Napa County', 'Nevada County', 'Orange County', 'Placer County',
'Plumas County', 'Riverside County', 'Sacramento County', 'San Benito County', 'San Bernardino County', 'San
Diego County', 'San Francisco County', 'San Joaquin County', 'San Luis Obispo County', 'San Mateo County', 'S
anta Barbara County', 'Santa Clara County', 'Santa Cruz County', 'Shasta County', 'Sierra County', 'Siskiyou
County', 'Solano County', 'Sonoma County', 'Stanislaus County', 'Sutter County', 'Tehama County', 'Trinity Co
unty', 'Tulare County', 'Tuolumne County', 'Ventura County', 'Yolo County', 'Yuba County']

Exploratory Visualizations: California Data

In [17]:
```
# Boxpot of trips greater than 500 miles in California
trip_data_clean_CA.boxplot(by ='Date',
                           column =['Number of Trips >=500'],
                           figsize=(15,9),
                           grid = False)
```

Out[17]: <matplotlib.axes._subplots.AxesSubplot at 0x11cab2940>

This boxplot looks slightly better than the last plot, and conveys the same overall trend, but there are still too many data to be illustrated effectively in a boxplot.
Next, we will try a few scatterplots.

```python
# Scatterplot of California trips above 500 miles

# Handle date time conversions between pandas and matplotlib
from pandas.plotting import register_matplotlib_converters
register_matplotlib_converters()

# Create figure and plot space
fig, ax = plt.subplots(figsize=(15, 9))

# Add x-axis and y-axis
ax.scatter(trip_data_clean_CA['Date'],
        trip_data_clean_CA['Number of Trips >=500'],
        color='purple', alpha=0.5)

plt.xlabel('Date', fontsize=15)
plt.ylabel('Number of Trips >=500 miles', fontsize=15)
plt.suptitle('Number of Trips Taken in California (>=500 miles)\nJanuary 2019 - October 2020', fontsize=20)

plt.show()
```
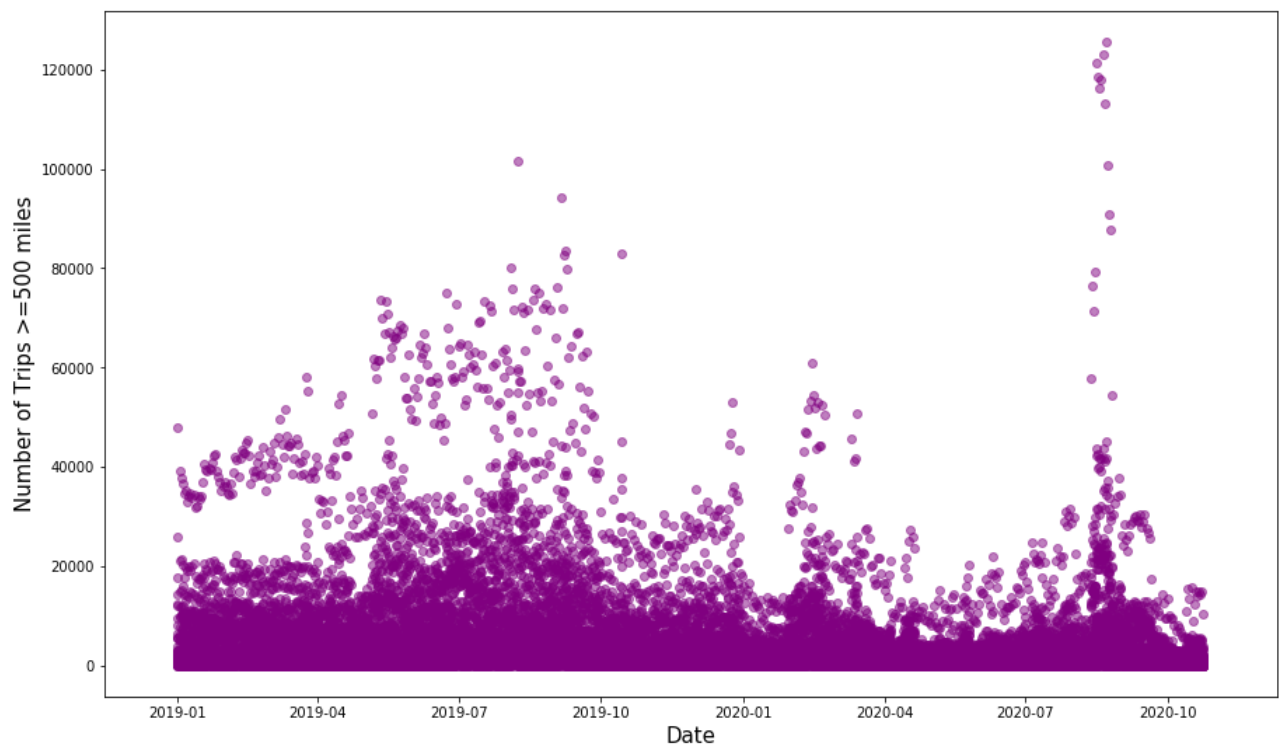

Number of Trips Taken in California (>=500 miles)
January 2019 - October 2020

```
# Scatterplot of California Population Staying at Home

# Create figure and plot space
fig, ax = plt.subplots(figsize=(15, 9))

# Add x-axis and y-axis
ax.scatter(trip_data_clean_CA['Date'],
        trip_data_clean_CA['Population Staying at Home'],
        color='blue', alpha=0.5)

plt.xlabel('Date', fontsize=15)
plt.ylabel('Population Staying at Home', fontsize=15)
plt.suptitle('Population Staying at Home in California\nJanuary 2019 - October 2020', fontsize=20)

plt.show()
```
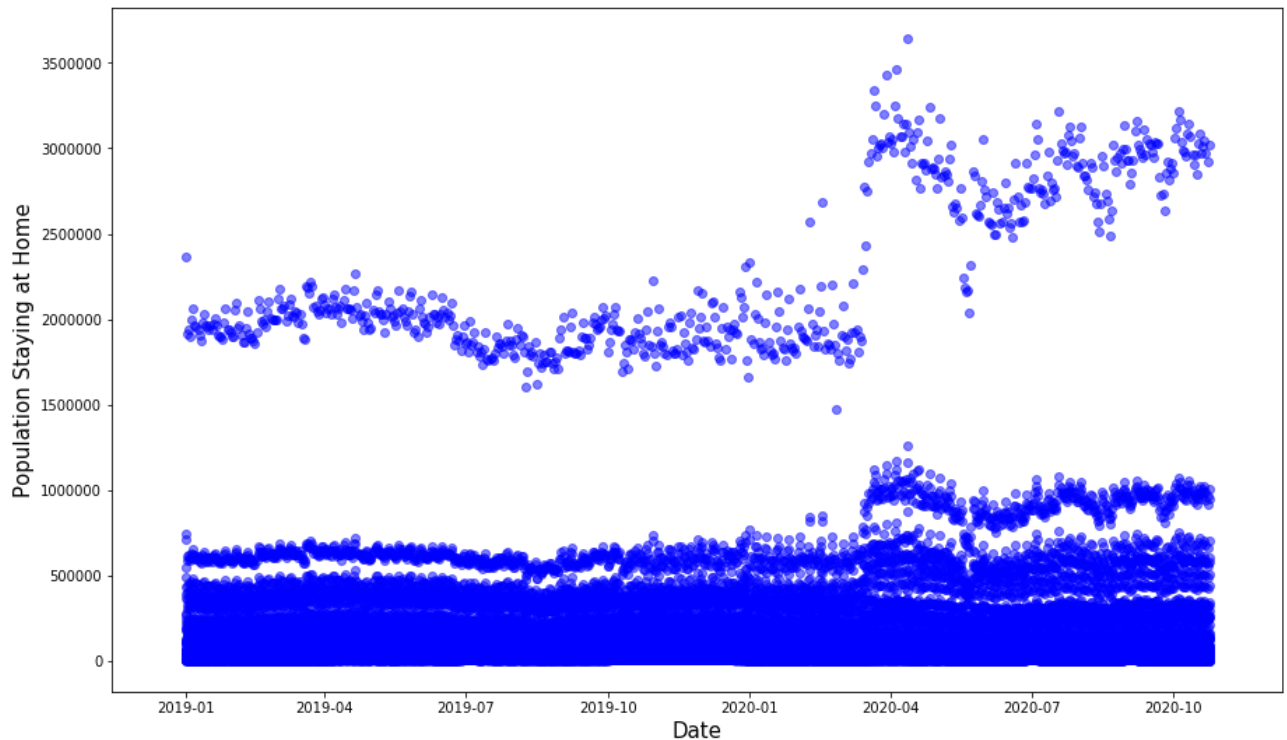

Population Staying at Home in California
January 2019 - October 2020

The first scatterplot depicting trips taken in California follows the expected pattern, while the second scatterplot depicting the population staying at home shows an opposite pattern but still expected. This makes sense that trips taken would be negatively correlated with population staying at home. In general, the less trips taken, the more people are staying at home.

To account for the dramatic change in trips taken and population staying at home beginning in March of 2020, it will be helpful to look at Covid-19 cases in California.

Covid-19 Supplemental Data

Data Source: https://data.ca.gov/dataset/covid-19-cases/resource/926fd08f-cc91-4828-af38-bd45de97f8c3 (https://data.ca.gov/dataset/covid-19-cases/resource/926fd08f-cc91-4828-af38-bd45de97f8c3)

Loading Data

```
In [20]:  # Loading data into a Pandas DataFrame
          covid_data_CA = pd.read_csv("statewide_cases.csv")

          # Checking dimensions
          print(covid_data_CA.shape)

          # Previewing data
          covid_data_CA.head(5)
```

```
(13925, 6)
```

Out[20]:

	county	totalcountconfirmed	totalcountdeaths	newcountconfirmed	newcountdeaths	date
0	Santa Clara	151.0	6.0	151	6	2020-03-18
1	Santa Clara	183.0	8.0	32	2	2020-03-19
2	Santa Clara	246.0	8.0	63	0	2020-03-20
3	Santa Clara	269.0	10.0	23	2	2020-03-21
4	Santa Clara	284.0	13.0	15	3	2020-03-22

Previously, our data included 11,225 rows, but now with updated data we have 13,925.

Cleaning Covid-19 Data

```
In [21]:  # Checking missing data sums
          covid_data_CA.isna().sum()
```

```
Out[21]:  county                 0
          totalcountconfirmed    3
          totalcountdeaths       2
          newcountconfirmed      0
          newcountdeaths         0
          date                   0
          dtype: int64
```

```
In [22]:  # Removing missing data
          covid_data_CA_clean = covid_data_CA.dropna()

          # Checking missing data sums
          covid_data_CA_clean.isna().sum()
```

```
Out[22]:  county                 0
          totalcountconfirmed    0
          totalcountdeaths       0
          newcountconfirmed      0
          newcountdeaths         0
          date                   0
          dtype: int64
```

```
In [23]:  covid_data_CA_clean.dtypes
```

```
Out[23]:  county                  object
          totalcountconfirmed    float64
          totalcountdeaths       float64
          newcountconfirmed        int64
          newcountdeaths           int64
          date                    object
          dtype: object
```

```
In [24]:  # Converting dates to datetime format
          covid_data_CA_clean['date'] = pd.to_datetime(covid_data_CA_clean['date'])

          covid_data_CA_clean['date'].head(10)
```

/Users/corinnemedeiros/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view
-versus-copy

```
Out[24]:  0    2020-03-18
          1    2020-03-19
          2    2020-03-20
          3    2020-03-21
          4    2020-03-22
          5    2020-03-23
          6    2020-03-24
          7    2020-03-25
          8    2020-03-26
          9    2020-03-27
          Name: date, dtype: datetime64[ns]
```

```
In [25]:  # Summary information for datetime column
          print("Datetime Data:\n")
          print(covid_data_CA_clean['date'].describe())
```

```
Datetime Data:

count                    13922
unique                     233
top        2020-09-01 00:00:00
freq                        60
first      2020-03-18 00:00:00
last       2020-11-05 00:00:00
Name: date, dtype: object
```

This updated data set includes up until November 5, 2020.
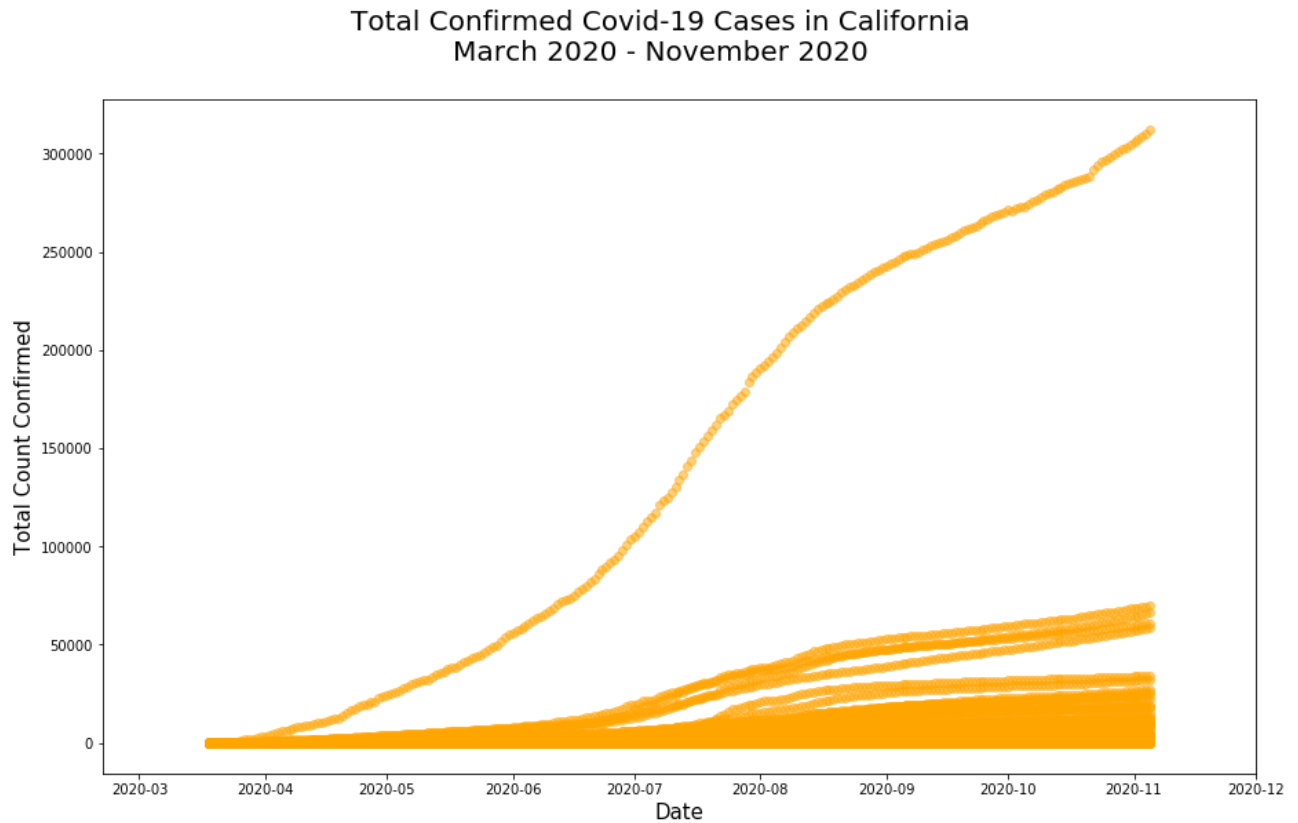
Plotting Covid-19 Data

`# Scatterplot of California Covid-19 cases`

```python
# Create figure and plot space
fig, ax = plt.subplots(figsize=(15, 9))

# Add x-axis and y-axis
ax.scatter(covid_data_CA_clean['date'],
        covid_data_CA_clean['totalcountconfirmed'],
        color='orange', alpha=0.5)

plt.xlabel('Date', fontsize=15)
plt.ylabel('Total Count Confirmed', fontsize=15)
plt.suptitle('Total Confirmed Covid-19 Cases in California\nMarch 2020 - November 2020', fontsize=20)

plt.show()
```

Total Confirmed Covid-19 Cases in California
March 2020 - November 2020

From this graph we can confirm that the amount of Covid-19 cases started rising in March of 2020, right as the number of trips taken started decreasing and the population staying at home started increasing. With the newly added data, we can see the recent even more dramatic rise in cases during our current month of November 2020.

At this point, we'll save our cleaned data as csv files to import into RStudio for modeling.

Saving Cleaned Data

In [29]:
```python
# Writing cleaned CA trip data to csv file
trip_data_clean_CA.to_csv('Trips_by_Distance_CA_clean.csv', index_label=False)
```

In [30]:
```python
# Writing cleaned CA Covid-19 data to csv file
covid_data_CA_clean.to_csv('covid_data_CA_clean.csv', index_label=False)
```