In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import plotly.plotly as py
import plotly.graph_objs as go
%matplotlib inline
```

In [2]:
```
# Read csv files
speed_10 = pd.read_csv("TX 2010 Speed Related Crashes Data.csv",low_memory=F
speed_11 = pd.read_csv("TX 2011 Speed Related Crashes Data.csv",low_memory=F
speed_12 = pd.read_csv("TX 2012 Speed Related Crashes Data.csv",low_memory=F
speed_13 = pd.read_csv("TX 2013 Speed Related Crashes Data.csv",low_memory=F
speed_14 = pd.read_csv("TX 2014 Speed Related Crashes Data.csv",low_memory=F
speed_15 = pd.read_csv("TX 2015 Speed Related Crashes Data.csv",low_memory=F
speed_16 = pd.read_csv("TX 2016 Speed Related Crashes Data.csv",low_memory=F
speed_17 = pd.read_csv("TX 2017 Speed Related Crashes Data.csv",low_memory=F
```

In [3]:
```
#Read the fips file for Texas and modified it
fips = pd.read_excel("Texas_FIPS.xlsx").astype(str)
county_fips = []
for county in fips['COUNTYFP']:
    if len(str(county)) == 1:
        county = str("00") + str(county)
        county_fips.append(county)
    elif len(str(county)) == 2:
        county = str("0") + str(county)
        county_fips.append(county)
    else:
        county = str(county)
        county_fips.append(county)
fips['COUNTYFP'] = county_fips
fips['FIPS'] = fips['STATEFP'] + fips['COUNTYFP']
fips.head()
```

Out[3]:

| | STATE | STATEFP | COUNTYFP | COUNTYNAME | CLASSFP | FIPS |
|---|---|---|---|---|---|---|
| **0** | TX | 48 | 001 | Anderson County | H1 | 48001 |
| **1** | TX | 48 | 003 | Andrews County | H1 | 48003 |
| **2** | TX | 48 | 005 | Angelina County | H1 | 48005 |
| **3** | TX | 48 | 007 | Aransas County | H1 | 48007 |
| **4** | TX | 48 | 009 | Archer County | H1 | 48009 |

In [20]:
```
def clean_data(csv_file):
    csv_file = csv_file[['Crash ID','Latitude', 'Longitude', 'City', 'County
    csv_file = csv_file[csv_file['Person Type'] == "Driver"]
    csv_file = csv_file[csv_file['Speed Limit'] >= 0]
    #Write the for loop to remove all those data that have "No Data" entry
    csv_file = csv_file.replace(to_replace = "No Data", value = np.nan)
    csv_file = csv_file.dropna(axis = 0, how = "any")
    return csv_file
```

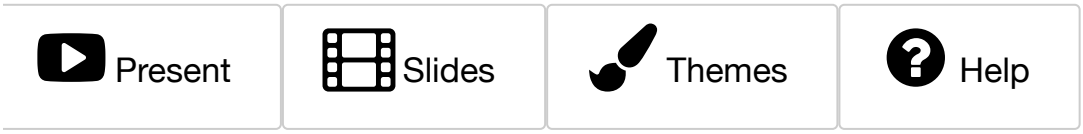
In [33]:
```
#Create a list contains all the file and reset the data with a clean data
csv_file = [speed_10, speed_11, speed_12, speed_13, speed_14, speed_15, spee
new_file = []
for file in csv_file:
    file = clean_data(file)
    county = []
    fips_code = []
    for element in file['County']:
        element = element + str(" County")
        county.append(element)
        # Make a new columns for the FIPS code
        if element in list(fips['COUNTYNAME']):
            i = (list(fips['COUNTYNAME']).index(element))
            fips_code.append(fips['FIPS'][i])
        else:
            fips_code.append(np.nan)
    file['County'] = county
    file['FIPS_CODE'] = fips_code
    #Drop NA again
    file = file.dropna(axis = 0, how = "any")
    new_file.append(file)
```

In [35]:
```
new_file[0].head()
```

Out[35]:

| | Crash ID | Latitude | Longitude | City | County | Crash Death Count | Crash Total Injury Count | Speed Limit | Weather Condition | Pe Ethn |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 11154479 | 33.66384399 | -95.53569239 | Paris | Lamar County | 0 | 1 | 30 | Cloudy | E |
| **1** | 11154479 | 33.66384399 | -95.53569239 | Paris | Lamar County | 0 | 1 | 30 | Cloudy | Hisp |
| **2** | 11154515 | 34.13157119 | -99.13679543 | Rural Wilbarger County | Wilbarger County | 0 | 0 | 70 | Cloudy | W |
| **5** | 11155308 | 27.58794537 | -99.52219213 | Laredo | Webb County | 0 | 1 | 30 | Clear | Hisp |
| **6** | 11155308 | 27.58794537 | -99.52219213 | Laredo | Webb County | 0 | 1 | 30 | Clear | Hisp |

In [ ]:
```
```


Present
Slides
Themes
Help