Bret Young

DSC 640

Assignment 6.2

19 November 2020

- histogram
- boxplot
- bullet chart

In [1]:
```
# Import required packages
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:
```
# Load dataset
url_1 = '~/Desktop/DSC 640/ex6-2/birth-rate.csv'
data_1 = pd.read_csv(url_1)
```

In [3]:
```
data_1.head()
```

Out[3]:

	Country	1960	1961	1962	1963	1964	1965	1966	1967	1968	...	19
0	Aruba	36.400	35.179	33.863	32.459	30.994	29.513	28.069	26.721	25.518	...	15.0
1	Afghanistan	52.201	52.206	52.208	52.204	52.192	52.168	52.130	52.076	52.006	...	51.2
2	Angola	54.432	54.394	54.317	54.199	54.040	53.836	53.585	53.296	52.984	...	48.6
3	Albania	40.886	40.312	39.604	38.792	37.913	37.008	36.112	35.245	34.421	...	17.7
4	Netherlands Antilles	32.321	30.987	29.618	28.229	26.849	25.518	24.280	23.173	22.230	...	15.8

5 rows × 50 columns

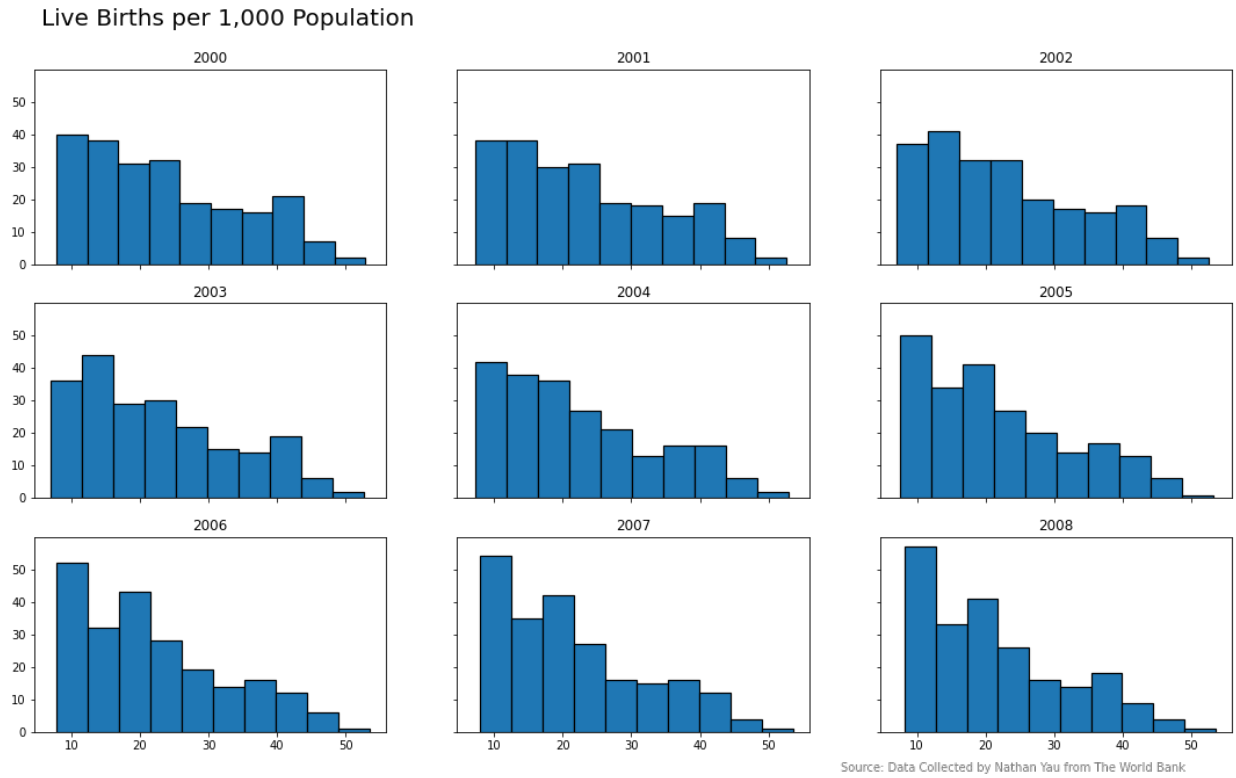
In [4]:
```
# select 2000+
data_1_filt = data_1.iloc[:,41:]

data_1_filt.head()
```

Out[4]:

	2000	2001	2002	2003	2004	2005	2006	2007	2008
0	14.528	14.041	13.579	13.153	12.772	12.441	12.159	11.919	11.716
1	50.903	50.486	49.984	49.416	48.803	48.177	47.575	47.023	46.538
2	48.355	48.005	47.545	46.936	46.184	45.330	44.444	43.607	42.875
3	16.850	16.081	15.444	14.962	14.644	14.485	14.464	14.534	14.649
4	15.412	15.096	14.824	14.565	14.309	14.051	13.790	13.532	13.281

In [5]:
```
# Create axes and figure
fig, (ax1, ax2, ax3) = plt.subplots(3, 3,sharex = True, sharey = True)

# Set figure size
fig.set_size_inches(18.5, 10.5)

# Add plots to figure
ax1[0].hist(data_1_filt['2000'], edgecolor = 'black', linewidth = 1.2)
ax1[1].hist(data_1_filt['2001'], edgecolor = 'black', linewidth = 1.2)
ax1[2].hist(data_1_filt['2002'], edgecolor = 'black', linewidth = 1.2)
ax2[0].hist(data_1_filt['2003'], edgecolor = 'black', linewidth = 1.2)
ax2[1].hist(data_1_filt['2004'], edgecolor = 'black', linewidth = 1.2)
ax2[2].hist(data_1_filt['2005'], edgecolor = 'black', linewidth = 1.2)
ax3[0].hist(data_1_filt['2006'], edgecolor = 'black', linewidth = 1.2)
ax3[1].hist(data_1_filt['2007'], edgecolor = 'black', linewidth = 1.2)
ax3[2].hist(data_1_filt['2008'], edgecolor = 'black', linewidth = 1.2)

# Set titles, caption and axis labels
fig.suptitle("Live Births per 1,000 Population", x = 0.25, y = 0.95, f
ontsize = 20)
fig.text(.87, .08, 'Source: Data Collected by Nathan Yau from The Worl
d Bank', ha = 'right', color = 'gray')

# add titles to subplots
ax1[0].set_title('2000')
ax1[1].set_title('2001')
ax1[2].set_title('2002')
ax2[0].set_title('2003')
ax2[1].set_title('2004')
ax2[2].set_title('2005')
ax3[0].set_title('2006')
ax3[1].set_title('2007')
ax3[2].set_title('2008')

# Show plot
plt.show

# save file
fig.savefig("python_histogram.png", bbox_inches = 'tight')
```

```
/Library/Frameworks/Python.framework/Versions/3.7/lib/python3.7/site
-packages/numpy/lib/histograms.py:839: RuntimeWarning: invalid value
encountered in greater_equal
  keep = (tmp_a >= first_edge)
/Library/Frameworks/Python.framework/Versions/3.7/lib/python3.7/site
-packages/numpy/lib/histograms.py:840: RuntimeWarning: invalid value
encountered in less_equal
  keep &= (tmp_a <= last_edge)
```



In [6]:
```
# Load dataset
url_2 = '~/Desktop/DSC 640/ex6-2/education.csv'
data_2 = pd.read_csv(url_2)
```

In [7]:
```
data_2.head()
```

Out[7]:

	state	reading	math	writing	percent_graduates_sat	pupil_staff_ratio	dropout_rate
0	United States	501	515	493	46	7.9	4.4
1	Alabama	557	552	549	7	6.7	2.3
2	Alaska	520	516	492	46	7.9	7.3
3	Arizona	516	521	497	26	10.4	7.6
4	Arkansas	572	572	556	5	6.8	4.6

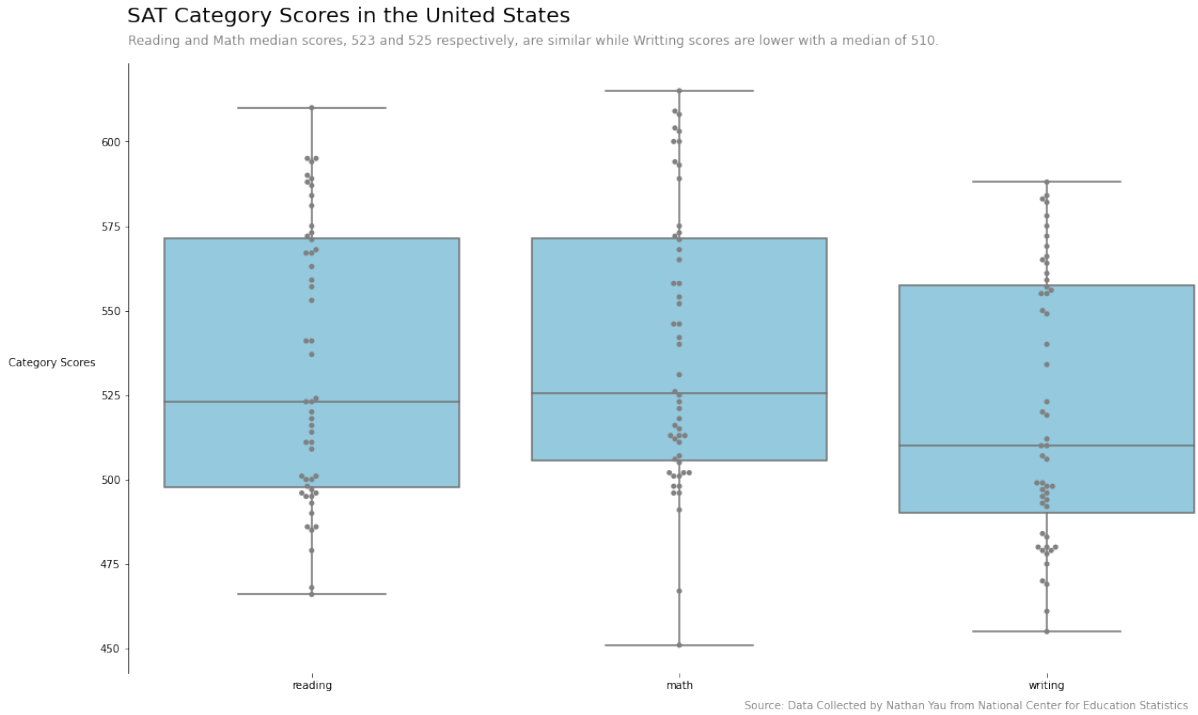
In [8]:
```
data_2_filt = pd.melt(data_2.iloc[:,1:4])

data_2_filt.head()
```

Out[8]:

	variable	value
0	reading	501
1	reading	557
2	reading	520
3	reading	516
4	reading	572

In [9]:
```
import seaborn as sns

# Create axes and figure
fig = plt.figure()
ax1 = fig.add_subplot(111)

# Set figure size
fig.set_size_inches(18.5, 10.5)

# Add plot to figure
sns.swarmplot(data = data_2_filt, x = "variable", y = "value", color="
grey")
sns.boxplot(data = data_2_filt, x = "variable", y = "value", color="sk
yblue")

# Set titles, caption and axis labels
fig.suptitle("SAT Category Scores in the United States", x = 0.28, y =
0.95, fontsize=20)
fig.text(.87, .08, 'Source: Data Collected by Nathan Yau from National
Center for Education Statistics', ha = 'right', color = 'gray')
ax1.set_title("Reading and Math median scores, 523 and 525 respectivel
y, are similar while Writting scores are lower with a median of 510.",
y = 1.02, loc='left', color = 'gray')
ax1.set_xlabel("")
ax1.set_ylabel("Category Scores", rotation = 0, ha = 'right')

# Remove frame
ax1.spines['right'].set_visible(False)
ax1.spines['top'].set_visible(False)
ax1.spines['bottom'].set_visible(False)

# Show plot
plt.show

# save file
fig.savefig("python_boxplot.png")
```

```
/Library/Frameworks/Python.framework/Versions/3.7/lib/python3.7/site
-packages/statsmodels/tools/_testing.py:19: FutureWarning: pandas.ut
il.testing is deprecated. Use the functions in the public API at pan
das.testing instead.
  import pandas.util.testing as tm
```



In [10]:
```
usa_data = data_2[data_2['state'] == 'United States']

usa_data['total_sat'] = (2/3) * (usa_data['reading'] + usa_data['math'
] + usa_data['writing'])

print(usa_data['total_sat'])

sat_min = 400
sat_max = 1600
sat_49 = 1050
sat_50 = 1070
sat_74 = 1200
sat_91 = 1350
```

```
0    1006.0
Name: total_sat, dtype: float64

/Library/Frameworks/Python.framework/Versions/3.7/lib/python3.7/site
-packages/ipykernel_launcher.py:3: 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: https://pandas.pydata.org/pand
as-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-co
py
  This is separate from the ipykernel package so we can avoid doing
imports until
```

In [33]:
```
# import library
import plotly.graph_objects as go

# create visual
fig = go.Figure(go.Indicator(
    mode = "number+gauge+delta", value = int(usa_data['total_sat']),
    domain = {'x': [0.1, 1], 'y': [0, 1]},
    title = {'text' :"<b>United States</b>", 'font': {"size": 18}},
    delta = {'reference': sat_50, 'position': "top"},
    gauge = {
        'shape': "bullet",
        'axis': {'range': [sat_min, sat_max]},
        'threshold': {
            'line': {'color': "red", 'width': 2},
            'thickness': 1,
            'value': sat_50},
        'steps': [
            {'range': [sat_min, sat_49], 'color': "gray"},
            {'range': [sat_49, sat_74], 'color': "darkgray"},
            {'range': [sat_74, sat_91], 'color': "lightgray"}],
        'bar': {'color': "blue"}}))

# add figure title, legend and set height
fig.update_layout(
    title = {
            'text': "United States SAT Scores",
            'y':0.9,
            'x':0.155},
    legend_title = "The target value is 1070 or thhe 50th percentile.
We can see the average for the United States falls below target.",
    font=dict(
            size = 18,
                ),
    height = 250,
    width = 1000,
    annotations = [dict(xref = 'paper',
                        yref = 'paper',
                        x = 0.09, y = 1.7,
                        showarrow = False,
                        text = 'This is my caption for the Plotly figu
re')])

fig.show()

# save image
fig.write_image('python_bullet.png')
```

In [ ]: