Week 1 & 2 Assignment

Name: Kesav Adithya Venkidusamy
Course: DSC640 - Data Presentation and Visualization
Instructor: Catherine Williams

These two weeks we are going to be focused on bar charts, stacked bar charts, pie charts, donut charts, and line charts using various tools to create these visualizations. Since the course is structured in 2-week blocks, every 2 weeks are going to be focused on a different visualization type that you create in various tools. This is to get you familiar with the pros-cons of Python, R, Tableau, and PowerBI – all tools you are going to run into in the real world, but that all have their own set of pros/cons. The book will cover one way to create a visualization, but we all know we have many tools at our disposal, so throughout these exercises you will be challenged to learn a variety of ways. Follow the recommendations and suggestions from your book on the design and use videos and other sources to find various ways to build the same visualizations.

You will be provided datasets, but you are welcome to use any data you like. You must consolidate all the charts into ONE document with each chart labeled with the type of chart and technology - for example: Python - Bar Chart. Failure to label and consolidate the charts will resort in points being taken off or a 0 for the assignment.

To get started, you are going to need to install some things, like R and Python (you likely already have these). You do not need to install all of these in Weeks 1 & 2 – you can decide to wait and install after you determine what you are going to use each week, but ultimately you will need at least one visualization tool like PowerBI or Tableau (can be something different if you have more experience) and then the ability to create visualizations using Python and R – you can pick which package you create them in. Not every visualization will be possible in one package, you will likely have to use multiple packages/libraries to get the job done.

1 bar chart, 1 stacked bar chart, 1 pie chart, 1 donut, and 1 line chart with Python

In [3]:
```python
## Importing libraries required for this exercise
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

1. Bar Chart

In [40]:
```python
## Load the dataset into dataframe
hotdog_df = pd.read_excel("hotdog-contest-winners.xlsm")
hotdog_df.head()
```

Out[40]:

	Year	Winner	Dogs eaten	Country	New record
0	1980	Paul Siederman & Joe Baldini	9.1	United States	0

	Year	Winner	Dogs eaten	Country	New record
1	1981	Thomas DeBerry	11.0	United States	0
2	1982	Steven Abrams	11.0	United States	0
3	1983	Luis Llamas	19.5	Mexico	0
4	1984	Birgit Felden	9.5	Germany	0

In [41]:
```python
## Update the column names to make it easier to use
hotdog_df.rename(columns = {'Dogs eaten':'Dogs_eaten'}, inplace = True)
hotdog_df.rename(columns = {'New record':'New_record'}, inplace = True)
hotdog_df.head()
```

Out[41]:

	Year	Winner	Dogs_eaten	Country	New_record
0	1980	Paul Siederman & Joe Baldini	9.1	United States	0
1	1981	Thomas DeBerry	11.0	United States	0
2	1982	Steven Abrams	11.0	United States	0
3	1983	Luis Llamas	19.5	Mexico	0
4	1984	Birgit Felden	9.5	Germany	0

In [42]:
```python
## Plotting bar chart for Year vs Dogs eaten
plt.figure(figsize=(15,10))
plt.bar(hotdog_df.Year, hotdog_df.Dogs_eaten, color = 'blue')
plt.xlabel("Year")
plt.ylabel("# of dogs eaten by winner")
plt.title("Number of hotdogs eaten by Year")
plt.show()
```

Number of hotdogs eaten by Year
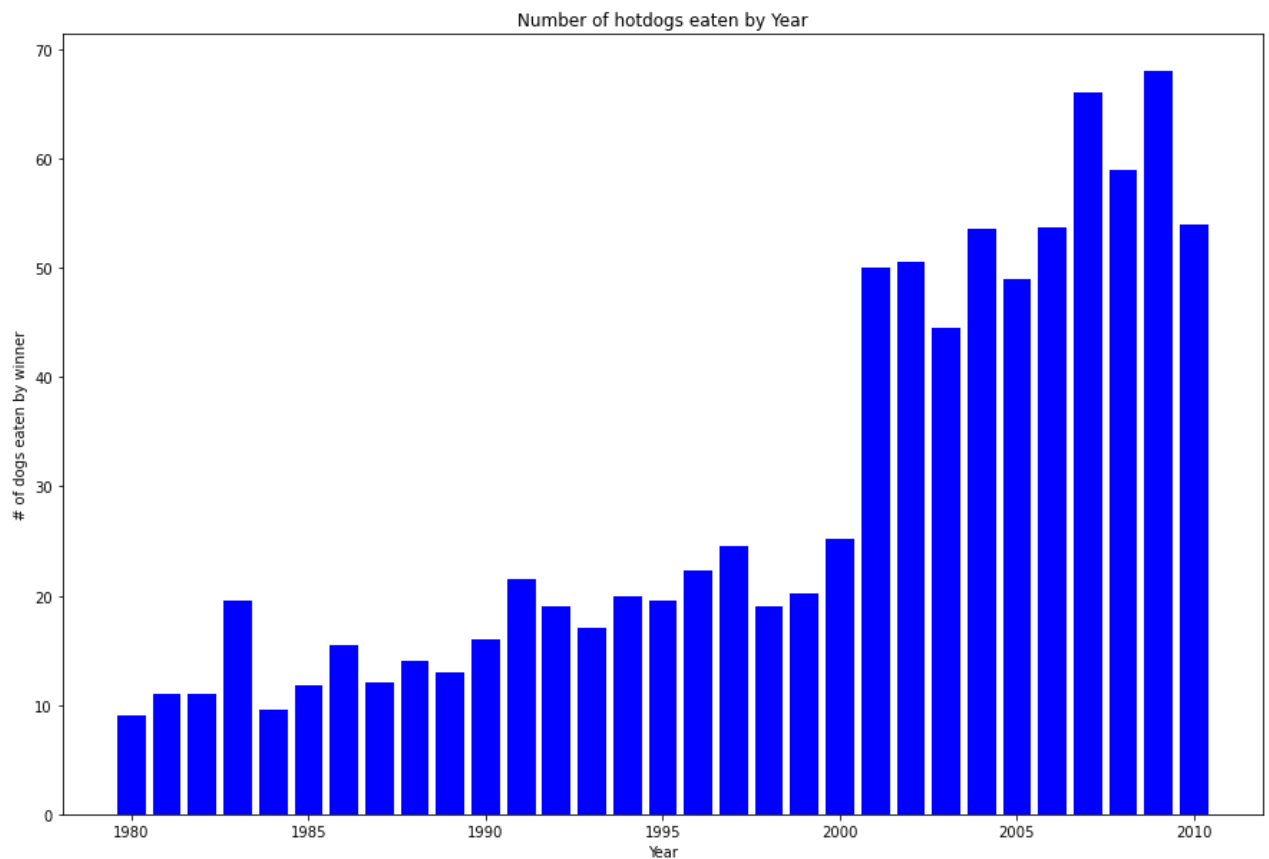
2. Stacked Bar Chart

```
In [24]:    # For this exercise I have considered the hotdog-places file
            hdp_df = pd.read_excel("hotdog-places.xlsm")
            hdp_df.head()
```

Out[24]:

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
0	25	50.0	50.5	44.5	53.5	49	54	66	59	68.0	54
1	24	31.0	26.0	30.5	38.0	37	52	63	59	64.5	43
2	22	23.5	25.5	29.5	32.0	32	37	49	42	55.0	37

```
In [25]:    ## Adding a field "Year" and place name for each row
            hdp_df['Year'] = ['Chicago','Texas','New_York']
            hdp_df = hdp_df.set_index('Year')
            hdp_df.head()
```

Out[25]:

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Year											
Chicago	25	50.0	50.5	44.5	53.5	49	54	66	59	68.0	54
Texas	24	31.0	26.0	30.5	38.0	37	52	63	59	64.5	43
New_York	22	23.5	25.5	29.5	32.0	32	37	49	42	55.0	37

```
In [26]:
```

```
## Taking transpose of the dataframe
##hdp_df = hdp_df.set_index('Year').T.rename_axis(None, 1).rename_axis('Year').reset_in
hdp_df = hdp_df.transpose()
hdp_df.reset_index()
hdp_df.head()
```

Out[26]:

Year	Chicago	Texas	New_York
2000	25.0	24.0	22.0
2001	50.0	31.0	23.5
2002	50.5	26.0	25.5
2003	44.5	30.5	29.5
2004	53.5	38.0	32.0

In [27]:
```
## Adding index as separate column
hdp_df['Year'] = hdp_df.index
hdp_df.head()
```

Out[27]:

Year	Chicago	Texas	New_York	Year
2000	25.0	24.0	22.0	2000
2001	50.0	31.0	23.5	2001
2002	50.5	26.0	25.5	2002
2003	44.5	30.5	29.5	2003
2004	53.5	38.0	32.0	2004

In [28]:
```
## Printng the colums and dtypes
hdp_df.dtypes
```

Out[28]:
```
Year
Chicago      float64
Texas        float64
New_York     float64
Year          object
dtype: object
```

In [38]:
```
## Plotting the stacked bar charts
plt.figure(figsize=(15,10))
plt.bar(hdp_df.Year, hdp_df.Chicago, label = 'Chicago', color = 'green')
plt.bar(hdp_df.Year, hdp_df.Texas, label = 'Texas', color = 'yellow')
plt.bar(hdp_df.Year, hdp_df.New_York, label = 'NewYork', color = 'red')
plt.xlabel("Year")
plt.ylabel("# of hotdog Restaurants")
plt.legend(loc = 'upper left')
plt.xticks(rotation = 90, fontsize = 10)
plt.title('Hotdog places growth chart by city')
plt.show()
```

Hotdog places growth chart by city

3. Pie Chart

In [44]:
```python
## We will use first dataframe to plot Pie chart
## Creating total number of Dogs eaten by country
hd_cnty_df = hotdog_df[['Dogs_eaten','Country']]
hd_cnty_df.head()
```

Out[44]:

	Dogs_eaten	Country
0	9.1	United States
1	11.0	United States
2	11.0	United States
3	19.5	Mexico
4	9.5	Germany

In [45]:
```python
## Calculate the total by country
hd_cnty_df2 = hd_cnty_df.groupby('Country').sum()
hd_cnty_df2.head()
```

Out[45]:

	Dogs_eaten
Country	
Germany	9.50
Japan	369.88

	Dogs_eaten
Country	
Mexico	19.50
United States	499.85

In [55]:
```python
## Plot pie chart based on the above calculated values
plt.figure(figsize=(15,10))
plt.pie(hd_cnty_df2.Dogs_eaten, labels = hd_cnty_df2.index, shadow=True)
plt.legend(title = "Country")
plt.title('Total Hotdog eaten by Country')
```

Out[55]:   Text(0.5, 1.0, 'Total Hotdog eaten by Country')

Total Hotdog eaten by Country
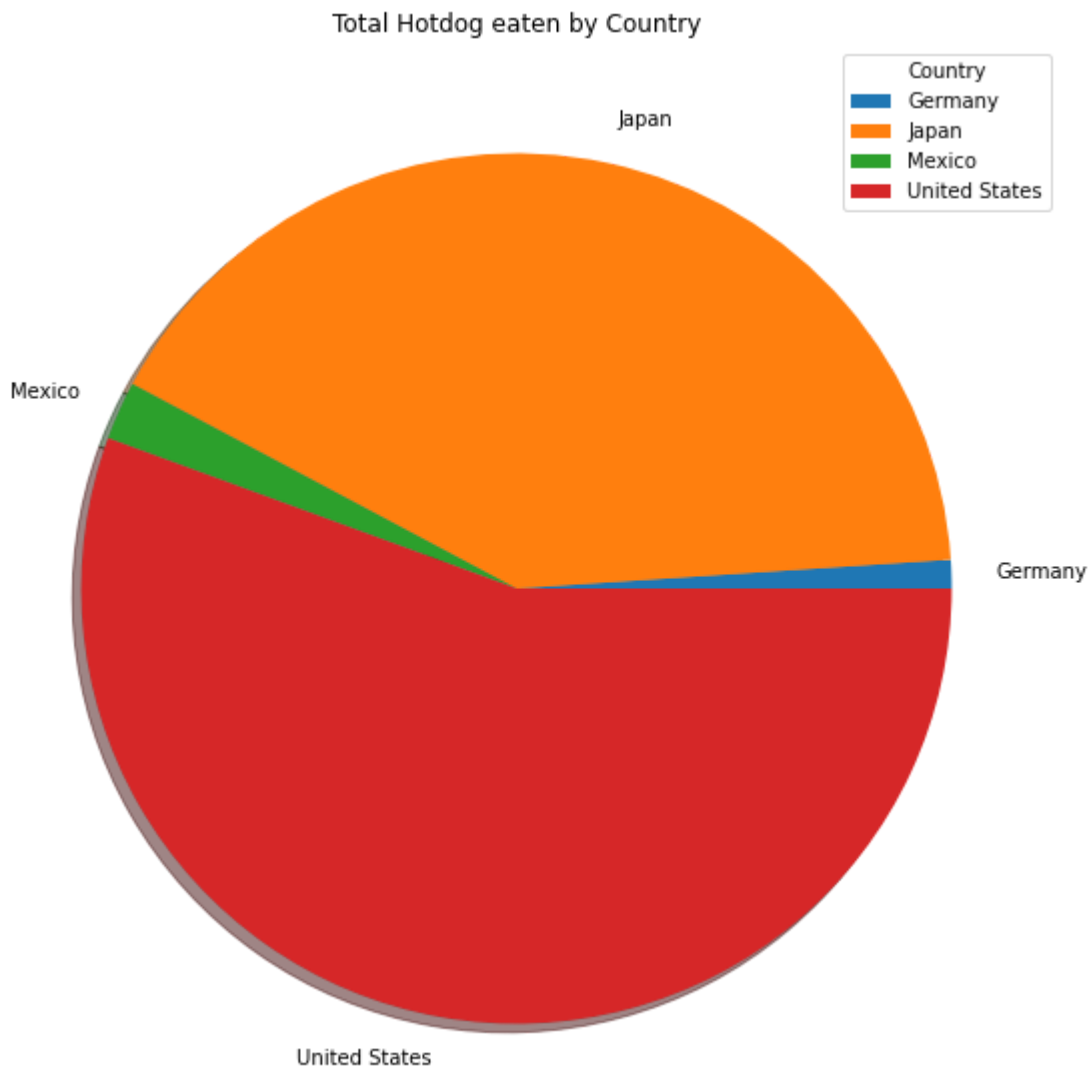
4. Donut Chart

For Donut Chart also, we will consider Hog eatens dataset; Here, I have created a sub dataframe with sum of Hog dog eaten by country

In [58]:
```python
## Plotting Donut chart
```

```
fig, ax1 = plt.subplots(figsize=(20, 15))
ax1.pie(hd_cnty_df2.Dogs_eaten, labels = hd_cnty_df2.index,shadow=True, startangle=90)

#draw a circle at the center of pie to make it look like a donut
centre_circle = plt.Circle((0,0),0.5,color='black', fc='white',linewidth=1.25)
fig = plt.gcf()
fig.gca().add_artist(centre_circle)
ax1.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.
plt.title('Total Hotdogs eaten by Country')
plt.legend(title = "Country")
plt.show()
```
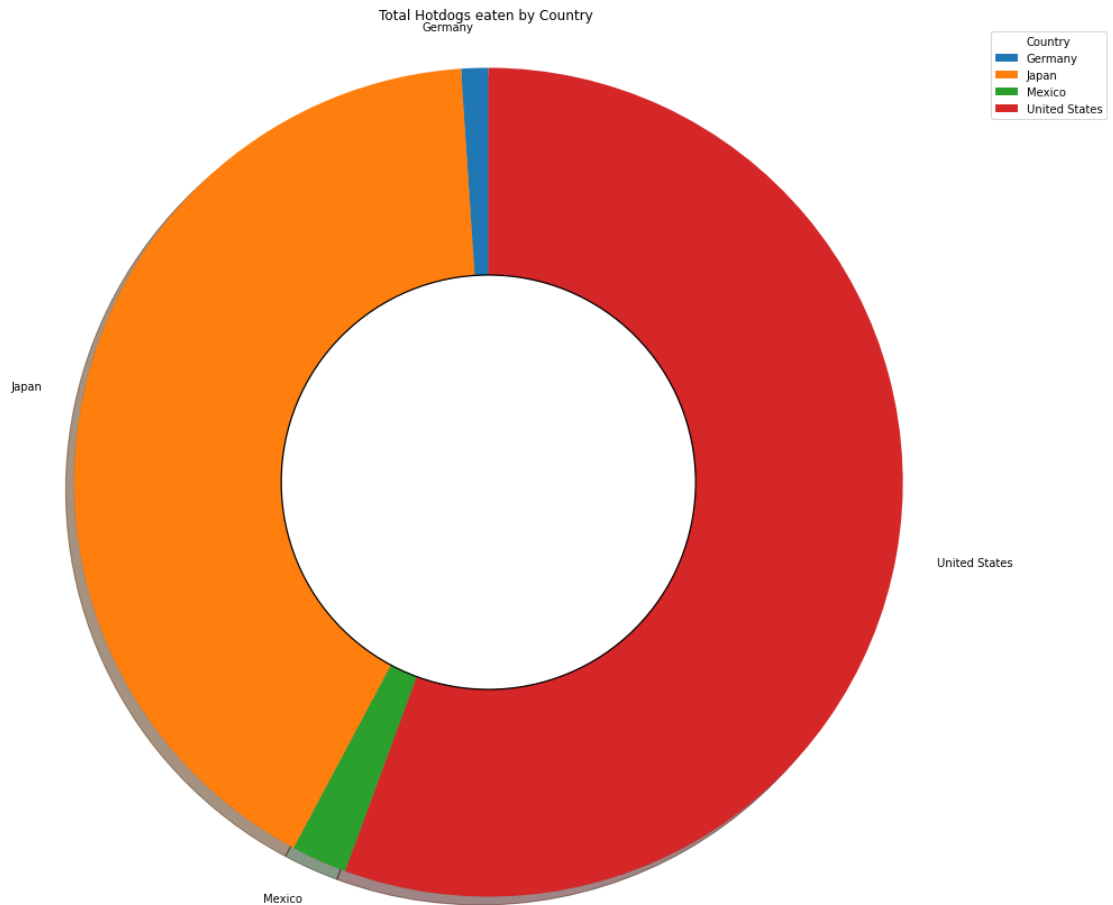

5. Line Chart

For this chart, we will consider the number of hot dogs eaten by year. From the dataset, we will consider the columns Year and number of hog dogs eaten.

In [68]:
```
# plot a line graph
plt.figure(figsize=(15, 10))
plt.plot(hotdog_df["Year"], hotdog_df["Dogs_eaten"])
plt.title('Hotdogs eaten by Year')
plt.xlabel("Year")
plt.ylabel("# of dogs eaten by winner")
plt.show()
```
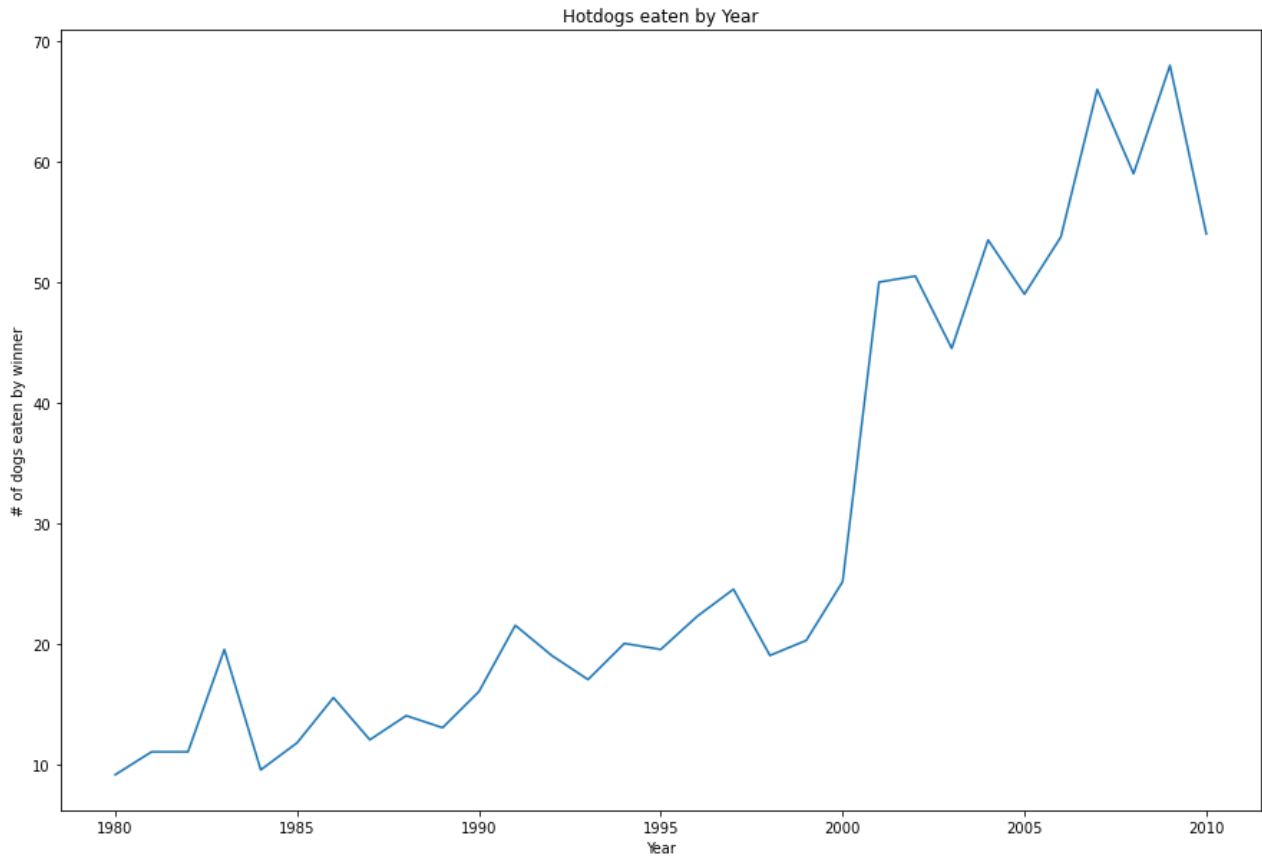
In [ ]:

Assignment_Week_1&2_Venkidusamy_KesavAdithya

Kesav Adithya Venkidusamy

2022/06/11

```r
knitr::opts_chunk$set(echo = TRUE)

library(readxl)
library(ggplot2)
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.1.3
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v tibble  3.1.4      v dplyr   1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.0.2      v forcats 0.5.1
## v purrr   0.3.4
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Data Loading

```r
hotdog_df <- read_excel("E:/Personal/Bellevue University/Course/github/dsc640/Week 1&2/hotdog-contest-w
head(hotdog_df)
```

```
## # A tibble: 6 x 5
##    Year Winner                    Dogs_eaten Country        'New record'
##   <dbl> <chr>                          <dbl> <chr>                 <dbl>
## 1  1980 Paul Siederman & Joe Baldini     9.1 United States             0
## 2  1981 Thomas DeBerry                  11   United States             0
## 3  1982 Steven Abrams                   11   United States             0
## 4  1983 Luis Llamas                     19.5 Mexico                    0
## 5  1984 Birgit Felden                    9.5 Germany                   0
## 6  1985 Oscar Rodriguez                 11.8 United States             0
```

```r
# Total number of records present in the data set
nrow(hotdog_df)
```

```
## [1] 31
```

```
## Create Bar Chart

ggplot(hotdog_df, aes(x=Country, y=Dogs_eaten)) +
  geom_bar(stat = "identity", fill = 'steelblue')+
  ggtitle("Number of Hotdogs Eaten by Country")+
  theme_minimal()
```

Number of Hotdogs Eaten by Country
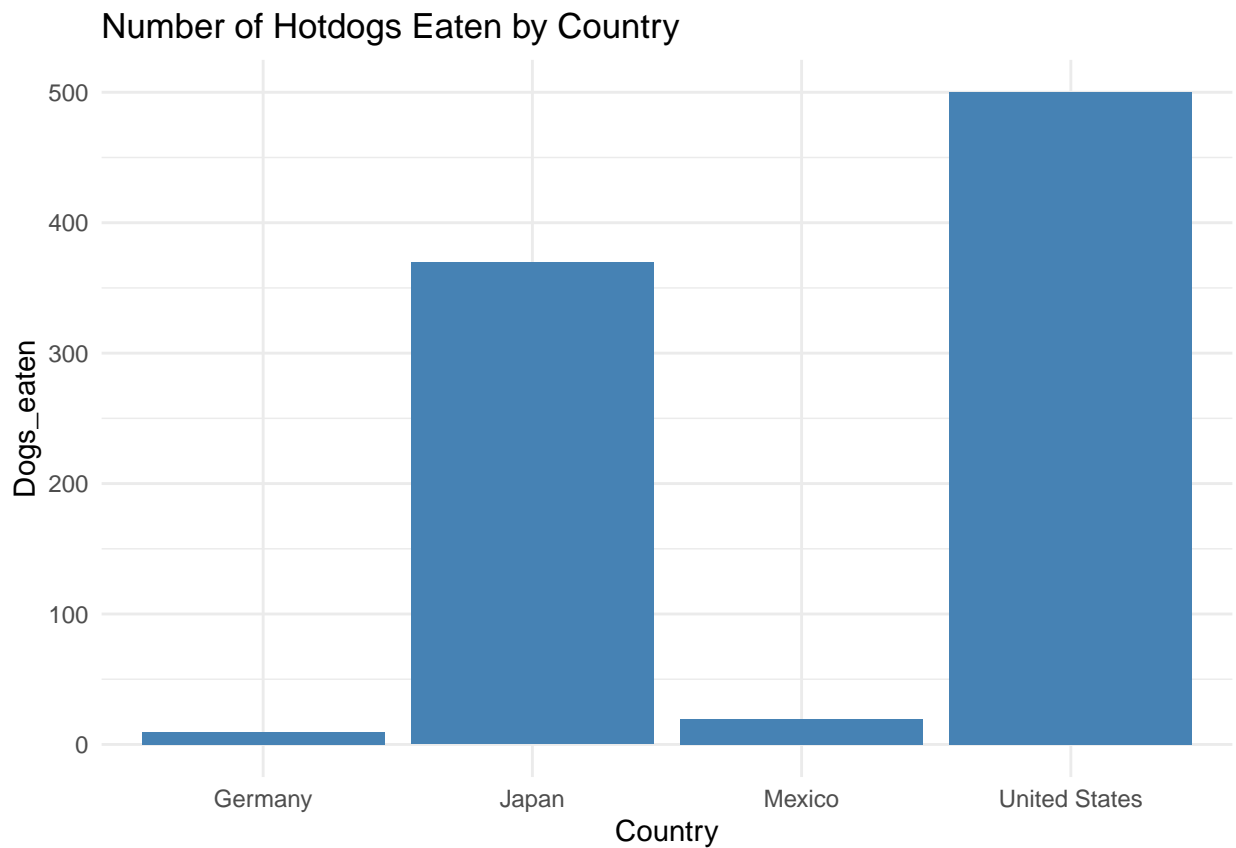

```
# Loadind the 2nd dataset for stacked bar
hdp_df <- read_excel("E:/Personal/Bellevue University/Course/github/dsc640/Week 1&2/hotdog-places.xlsm")

# Showing sample records
head(hdp_df)
```

```
## # A tibble: 3 x 11
##    '2000' '2001' '2002' '2003' '2004' '2005' '2006' '2007' '2008' '2009' '2010'
##     <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
## 1      25     50   50.5   44.5   53.5     49     54     66     59     68     54
## 2      24     31     26   30.5     38     37     52     63     59   64.5     43
## 3      22   23.5   25.5   29.5     32     32     37     49     42     55     37
```

```
## Plotting stacked bar plot
barplot(as.matrix(hdp_df),  main="# of Hot Dog Restaurant by Places",
        xlab="Year",
```

```
        ylab="# of Hot Dog Restaurants",
        legend.text =  c('Chicago','Texas','NY'),
        args.legend = list(x = "topright",bty='n', inset=c(-0.1,0))
        )
```

of Hot Dog Restaurant by Places
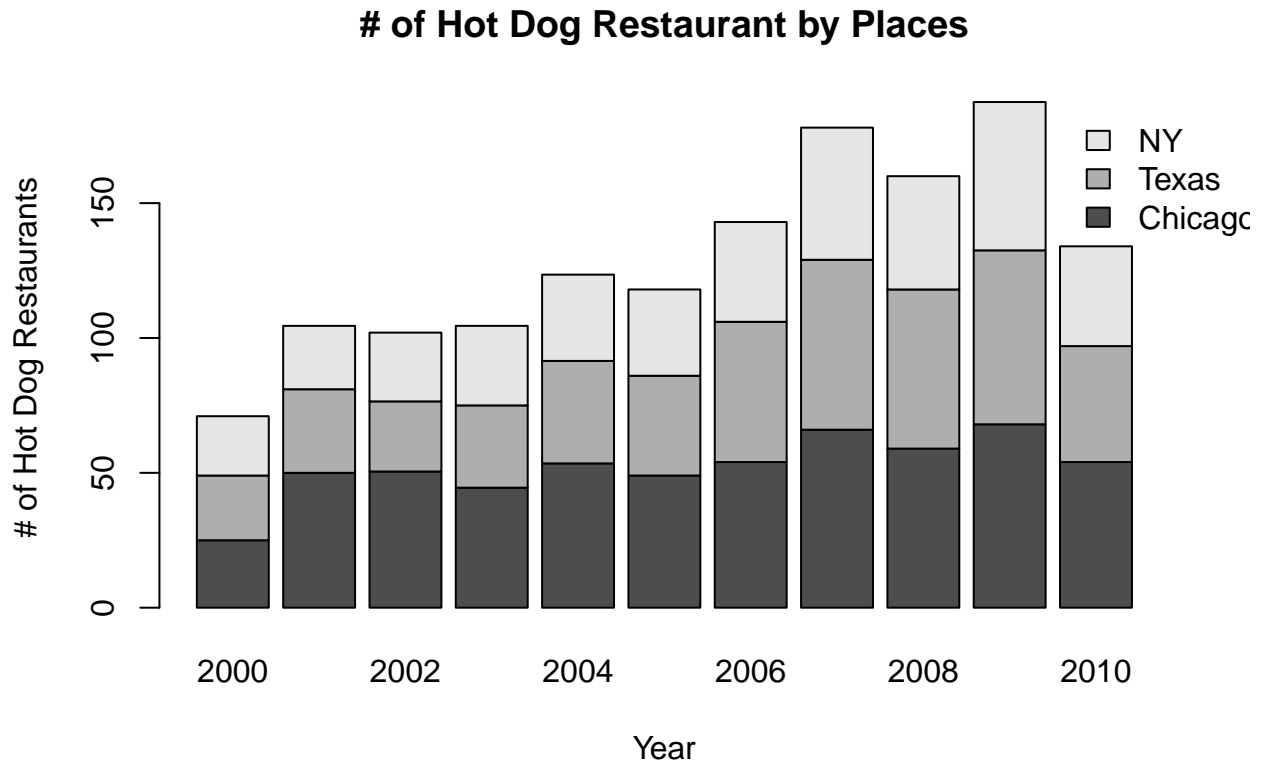

```
#Read in third file for the pie and donut charts
obama <- read_excel('E:/Personal/Bellevue University/Course/github/dsc640/Week 1&2/obama-approval-ratin

## Displaying few records
head(obama)
```

```
## # A tibble: 6 x 4
##   Issue          Approve Disapprove  None
##   <chr>            <dbl>      <dbl> <dbl>
## 1 Race Relations      52         38    10
## 2 Education           49         40    11
## 3 Terrorism           48         45     7
## 4 Energy Policy       47         42    11
## 5 Foreign Affairs     44         48     8
## 6 Environment         43         51     6
```

```
#Set up for pie and donut charts
#Creating smaller dataframe from original dataset
dat <- data.frame(count=c(obama$Approve), category=c(obama$Issue))
```

```
#Compute percentages
dat$fraction = dat$count / sum(dat$count)

#Compute the cumulative percentages
dat$ymax = cumsum(dat$fraction)

#Compute the bottom of each rectangle
dat$ymin = c(0, head(dat$ymax, n=-1))

#Compute label position
dat$labelPosition <- (dat$ymax + dat$ymin) / 2

#Compute a good label
dat$label <- paste0(dat$count)

#Pie chart
ggplot(dat, aes(ymax=ymax, ymin=ymin, xmax=4, xmin=3, fill=category)) +
  geom_rect() +
  geom_label( x=4, aes(y=labelPosition, label=label), size=3) +
  ggtitle("Approval Ratings by Issue") +
  coord_polar(theta="y") +
  theme_void()
```
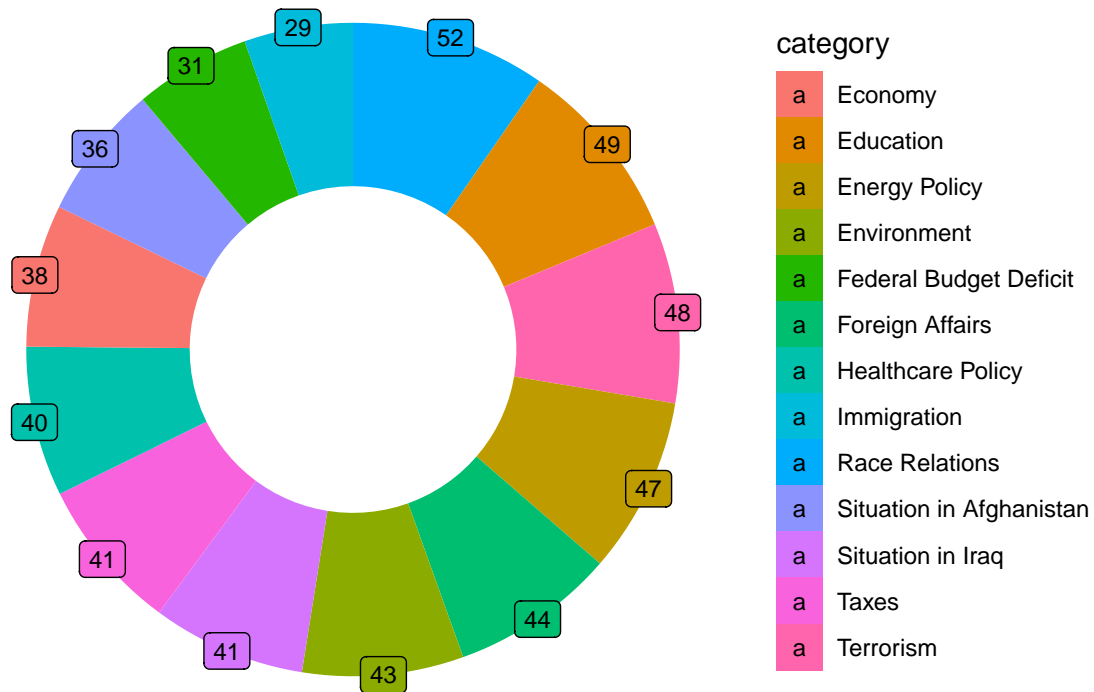
Approval Ratings by Issue
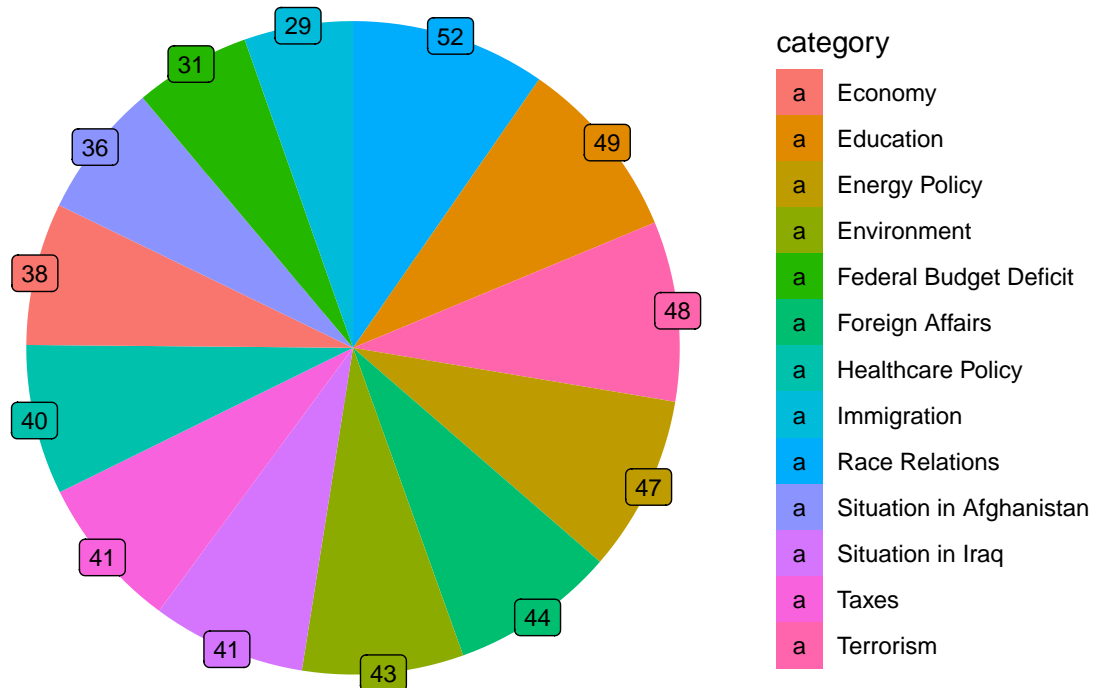

```
#Donut chart
ggplot(dat, aes(ymax=ymax, ymin=ymin, xmax=4, xmin=3, fill=category)) +
```

```
geom_rect() +
geom_label( x=4, aes(y=labelPosition, label=label), size=3) +
ggtitle("Approval Ratings by Issue") +
coord_polar(theta="y") +
theme_void() +
xlim(c(2, 4))
```

Approval Ratings by Issue


```
# We will use the same hotdog dataframe for this plot

ggplot(hotdog_df, aes(x=Year, y=Dogs_eaten)) +
  geom_line() +
  ggtitle("Number of Hotdogs Eaten by Winner by Year")
```
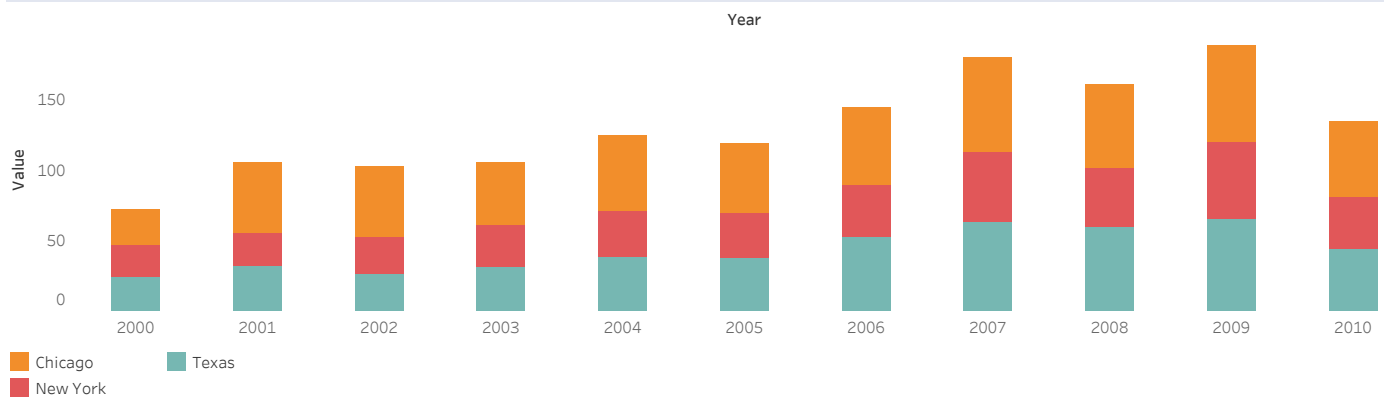
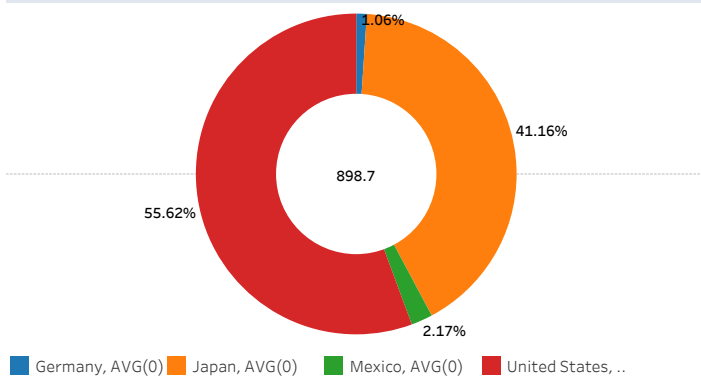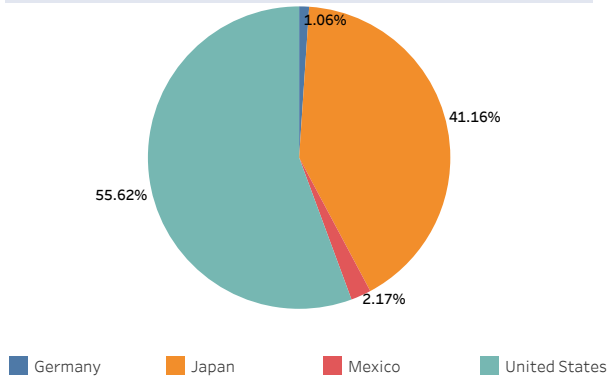Number of Hotdogs Eaten by Winner by Year
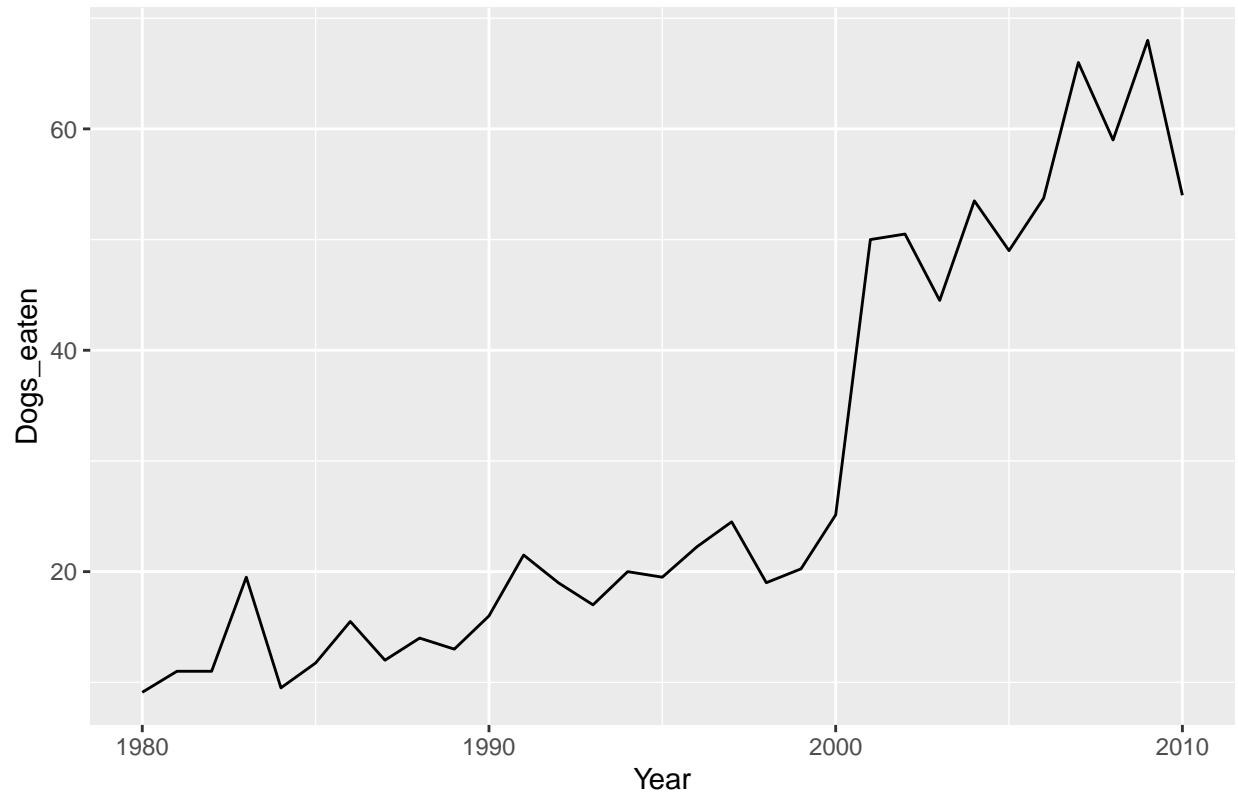
DSC640 Week 1&2 Assignment

of Hot Dogs Eaten by Year

Year vs Hog Dog Restaurant Regions

- Chicago
- Texas
- New York

Country vs # of Hot Dogs Eaten

898.7

1.06%
41.16%
55.62%
2.17%

- Germany, AVG(0)
- Japan, AVG(0)
- Mexico, AVG(0)
- United States, ..

Country vs # of Hot Dogs Eaten

1.06%
41.16%
55.62%
2.17%

- Germany
- Japan
- Mexico
- United States

of Hot Dogs eaten for each Year

of Hot Dogs Eaten by Year
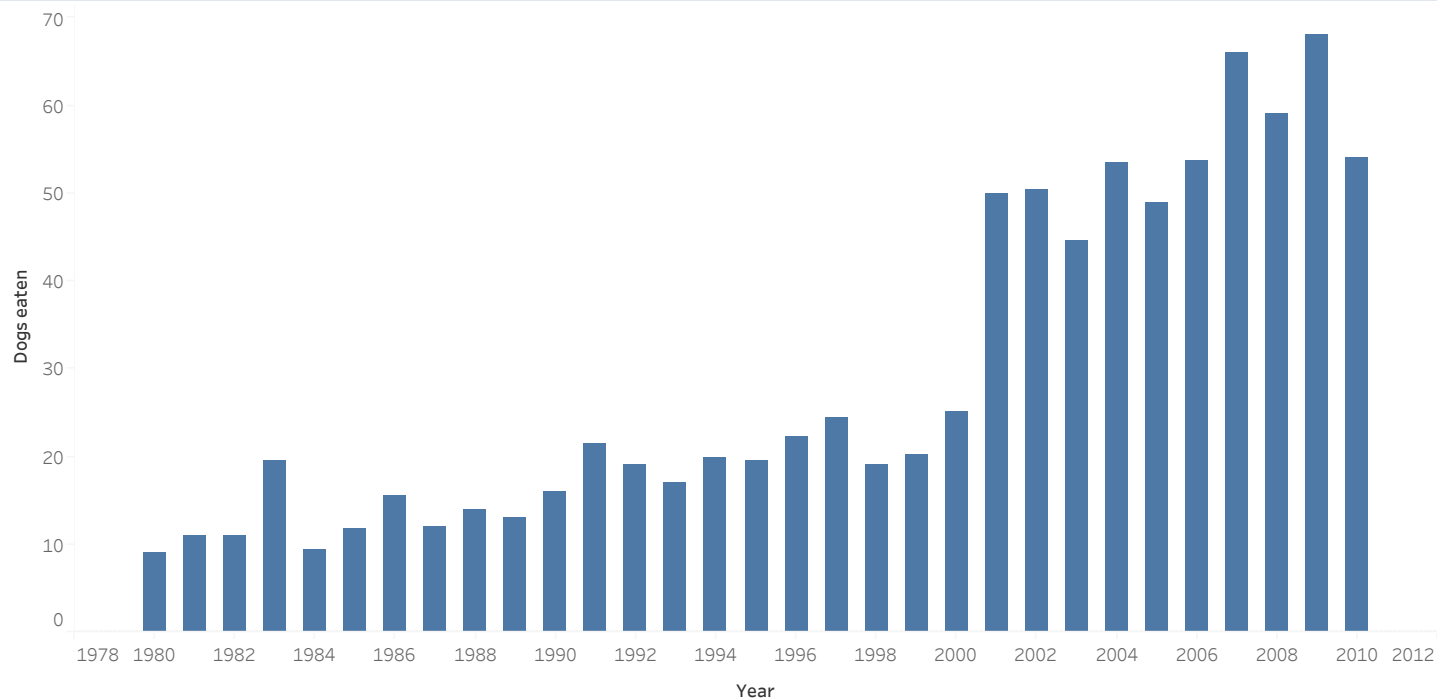
The plot of sum of Dogs eaten for Year.

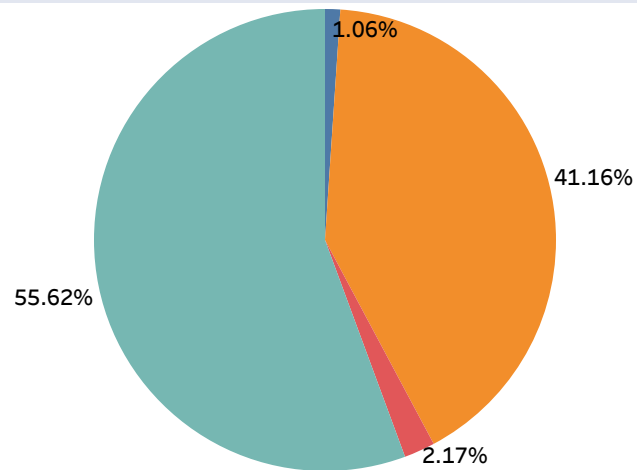# Year vs Hog Dog Restaurant Regions
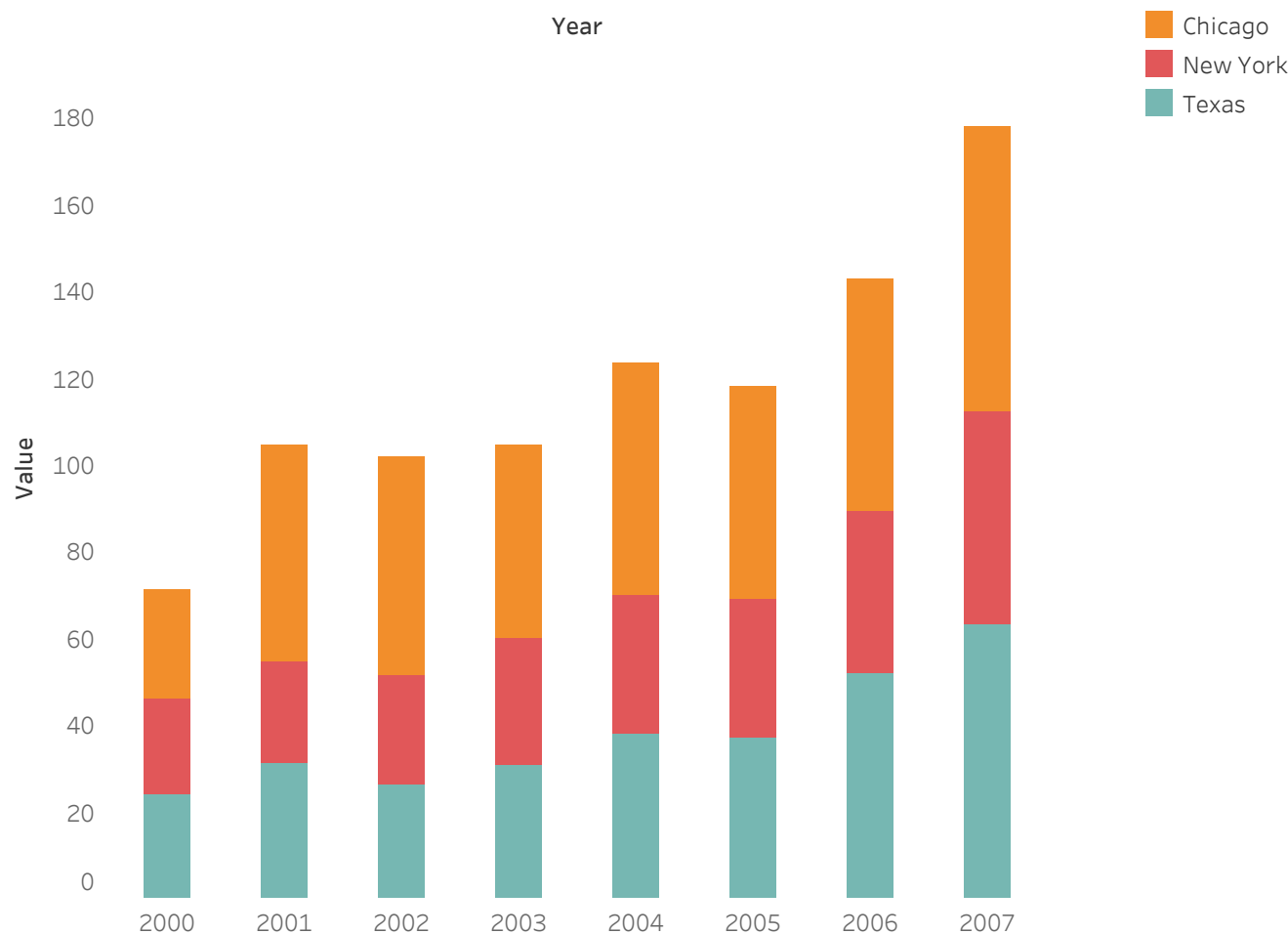
Year

Chicago, New York and Texas for each Year. Color shows details about Chicago, New York and Texas.

Year vs Hog Dog Restaurant Regions
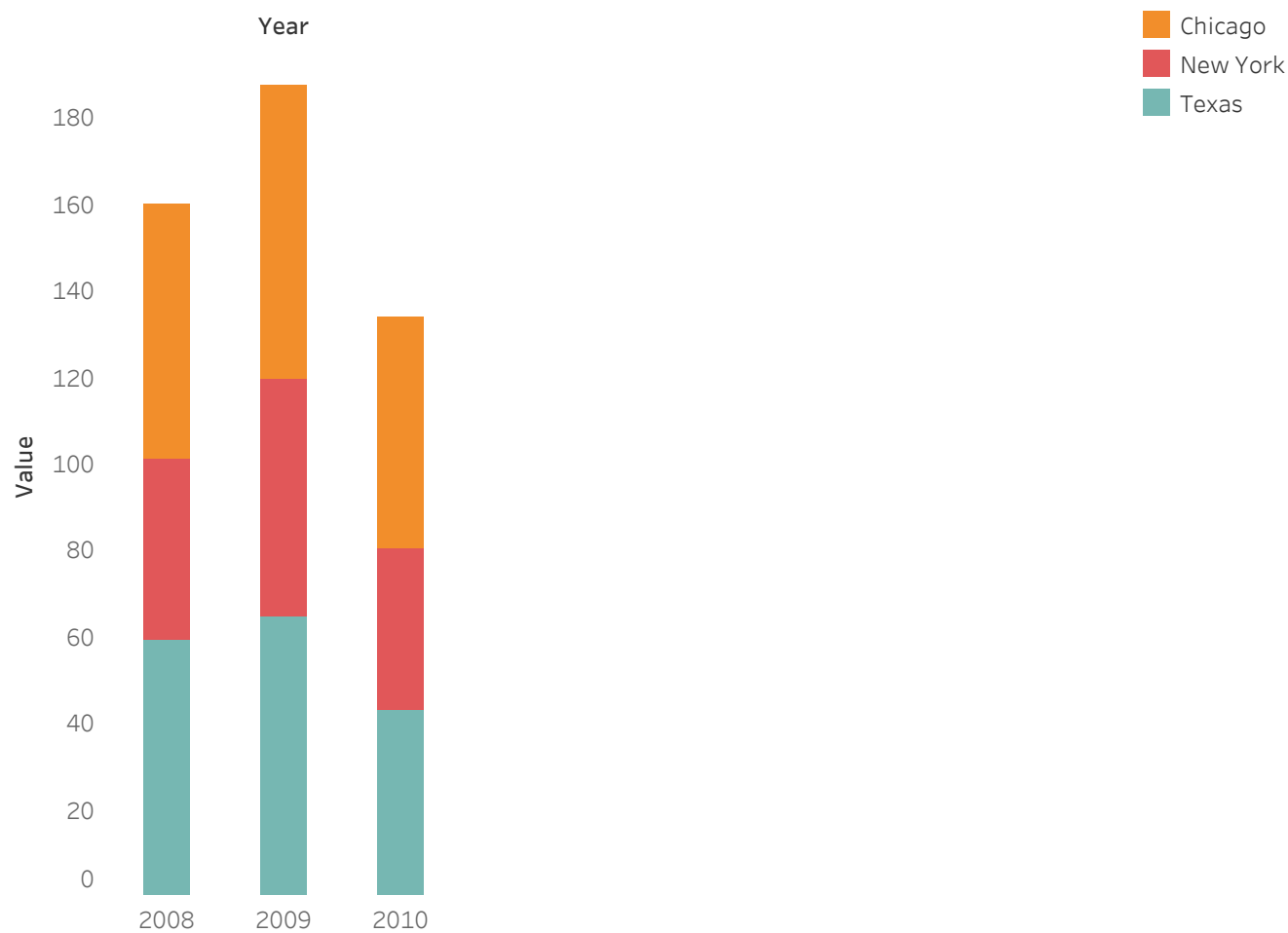
Year

Chicago, New York and Texas for each Year. Color shows details about Chicago, New York and Texas.
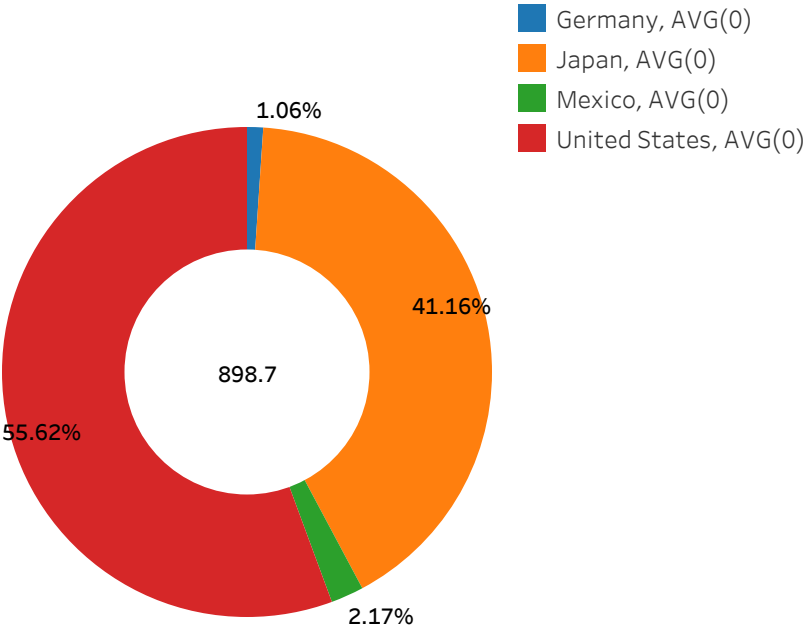
Country vs # of Hot Dogs Eaten

Country (color) and sum of Dogs eaten (size).

1.06%

41.16%

55.62%

2.17%

898.7

- Germany
- Japan
- Mexico
- United States

Country vs # of Hot Dogs Eaten

1.06%

41.16%

898.7

55.62%

2.17%

AVG(0) and AVG(0). For pane AVG(0):
Color shows details about Country and
AVG(0). For pane AVG(0): The marks are
labeled by sum of Dogs eaten.

of Hot Dogs eaten for each Year
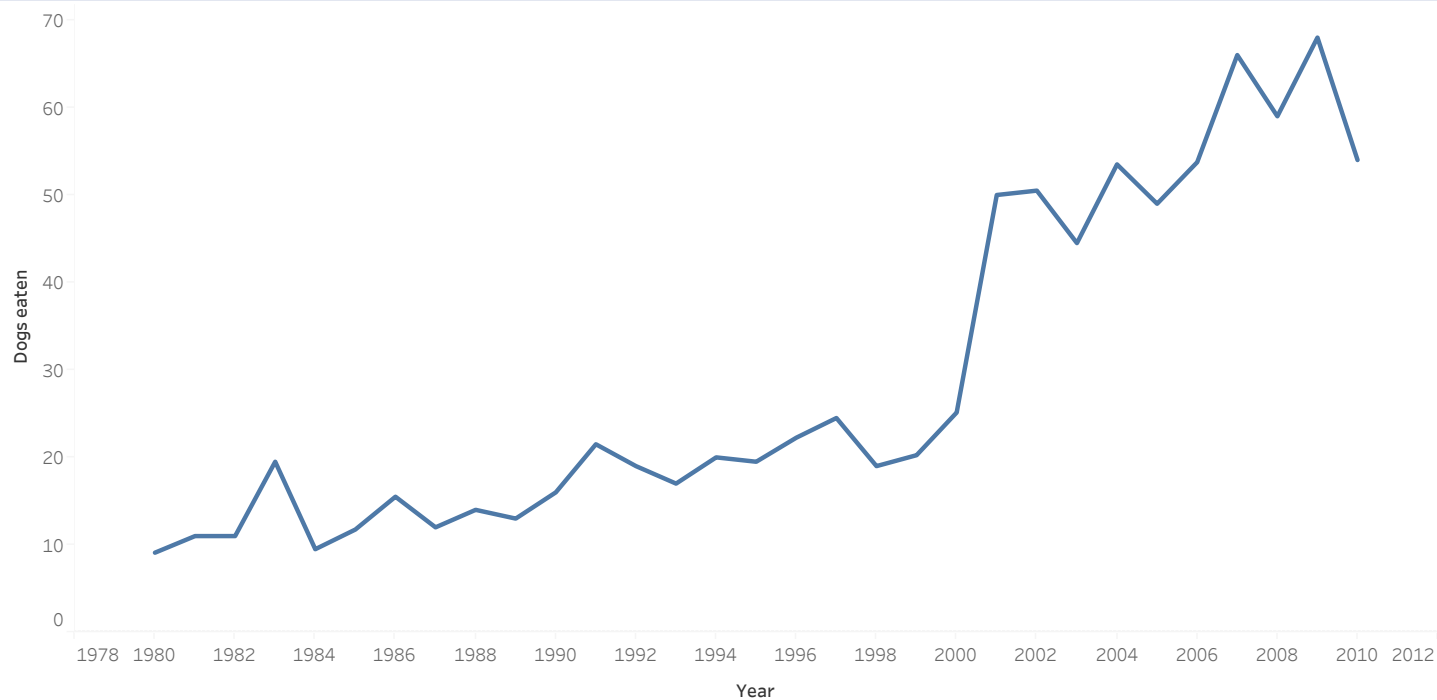
The trend of sum of Dogs eaten for Year.