**Weeks 1 & 2**

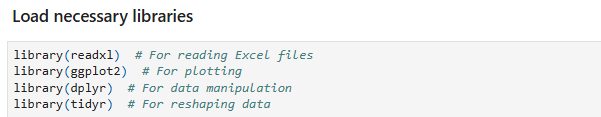
**McKenzie Payne**

**Assignment 1.2: Charts**

**Black Board Instructions:**

* **1 bar chart, 1 stacked bar chart, 1 pie chart, 1 donut, and 1 line chart with Tableau or PowerBI**
* **1 bar chart, 1 stacked bar chart, 1 pie chart, 1 donut, and 1 line chart with Python**
* **1 bar chart, 1 stacked bar chart, 1 pie chart, 1 donut, and 1 line chart with R**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Using R:**



**Paths to the data files**

In [2]:

obama\_approval\_path **<-** "C:/Users/mcken/Downloads/obama-approval-ratings.xls"

hotdog\_places\_path **<-** "C:/Users/mcken/Downloads/hotdog-places.xlsm"

hotdog\_winners\_path **<-** "C:/Users/mcken/Downloads/hotdog-contest-winners.xlsm"

**Load the data into R data frames**

In [3]:

obama\_approval **<-** read\_excel(obama\_approval\_path)

hotdog\_places **<-** read\_excel(hotdog\_places\_path)

hotdog\_winners **<-** read\_excel(hotdog\_winners\_path)

**Display the first few rows of each dataset**

In [4]:

cat("Obama Approval Ratings Data:\n")

print(head(obama\_approval))

cat("\nHotdog Places Data:\n")

print(head(hotdog\_places))

cat("\nHotdog Winners Data:\n")

print(head(hotdog\_winners))

Obama Approval Ratings Data:

# A tibble: 6 × 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

Hotdog Places Data:

# A tibble: 3 × 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

Hotdog Winners Data:

# A tibble: 6 × 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

**Bar Chart**

In [5]:

*# Reshape the hotdog\_places data from wide to long format*

hotdog\_places\_long **<-** hotdog\_places **%>%**

pivot\_longer(cols **=** everything(), names\_to **=** "Year", values\_to **=** "HotdogsEaten")

*# Create bar plot*

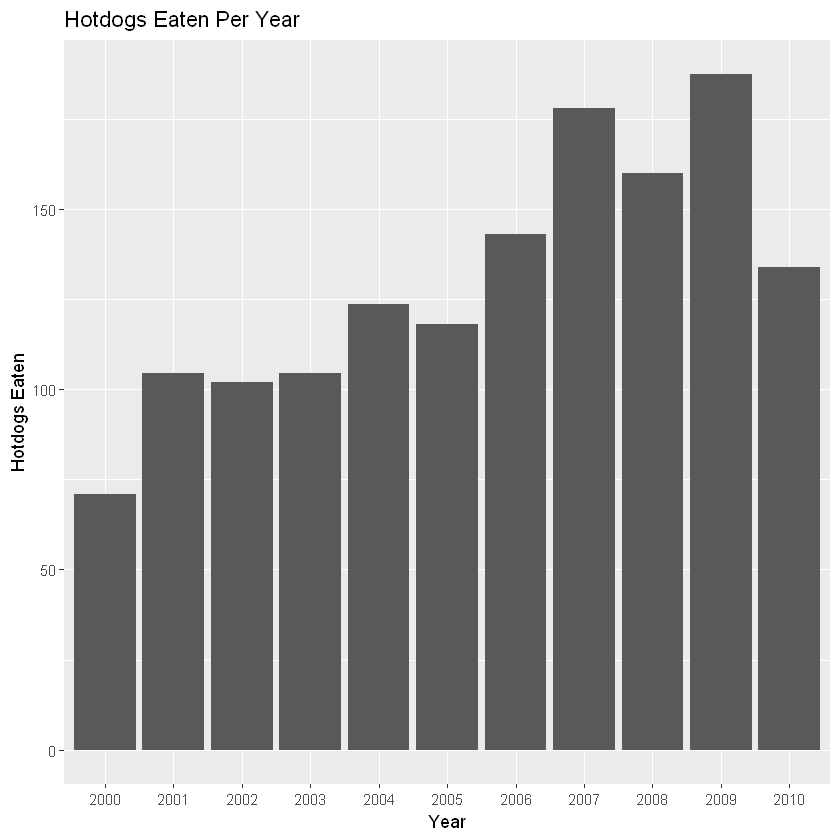
ggplot(hotdog\_places\_long, aes(x **=** Year, y **=** HotdogsEaten)) **+**

geom\_bar(stat **=** "identity") **+**

ggtitle('Hotdogs Eaten Per Year') **+** *# Set title*

xlab('Year') **+** *# Set x-axis label*

ylab('Hotdogs Eaten') *# Set y-axis label*



**Stacked Bar Chart**

In [6]:

*# Add dummy 'BunsEaten' data*

hotdog\_places\_long **<-** hotdog\_places\_long **%>%**

mutate(BunsEaten **=** HotdogsEaten **\*** 0.8)

**Create stacked bar plot**

In [7]:

ggplot(hotdog\_places\_long, aes(x **=** Year, y **=** HotdogsEaten, fill **=** "HotdogsEaten")) **+**

geom\_bar(stat **=** "identity") **+**

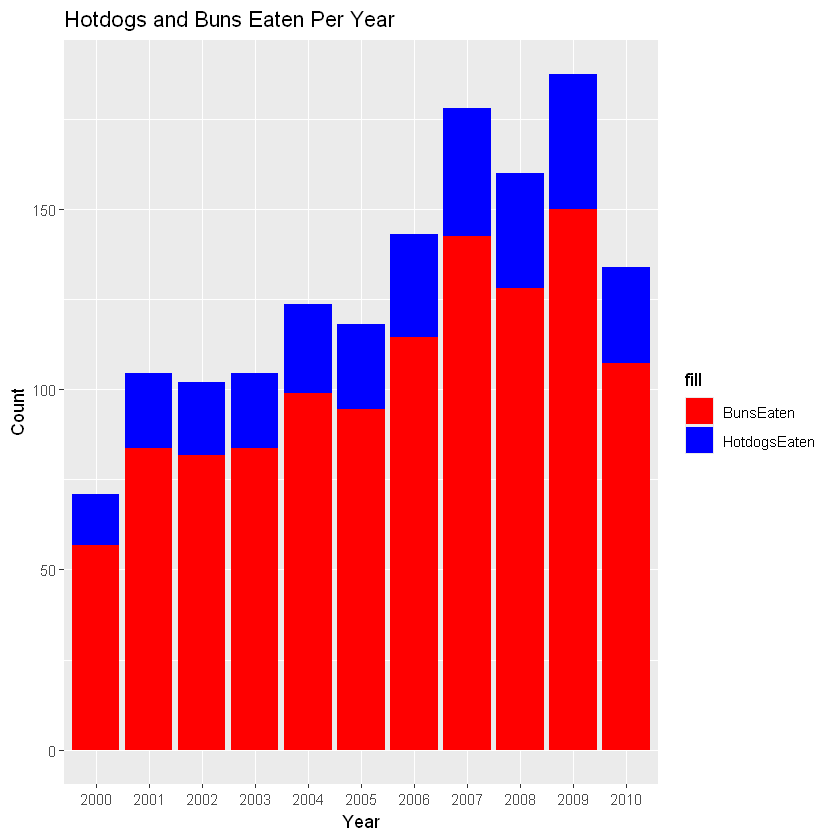
geom\_bar(aes(y **=** BunsEaten, fill **=** "BunsEaten"), stat **=** "identity") **+**

ggtitle('Hotdogs and Buns Eaten Per Year') **+** *# Set title*

xlab('Year') **+** *# Set x-axis label*

ylab('Count') **+** *# Set y-axis label*

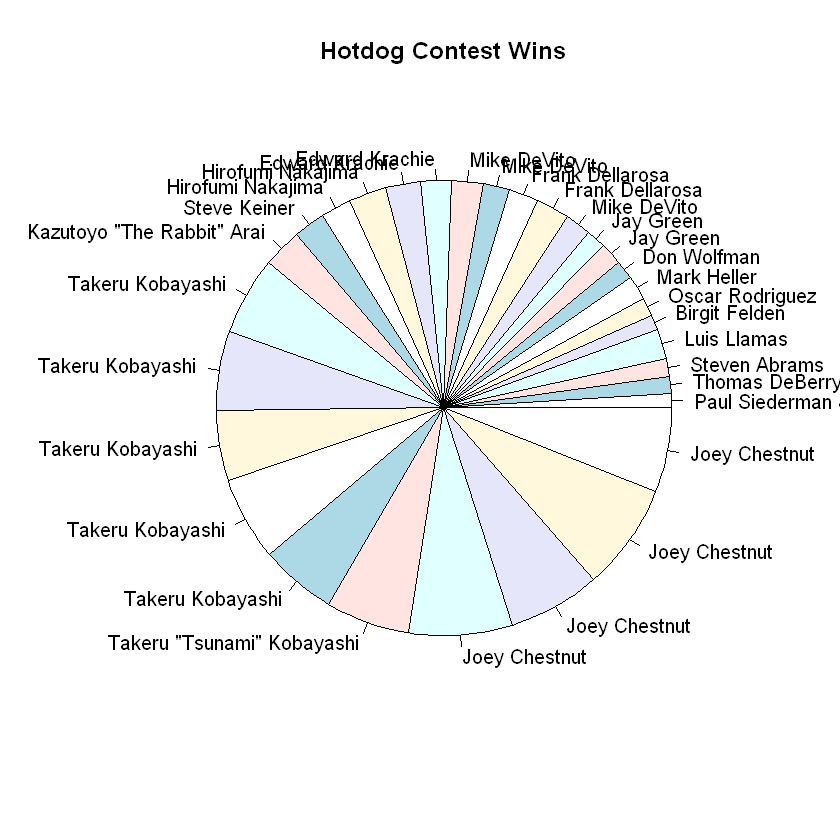
scale\_fill\_manual(values **=** c("HotdogsEaten" **=** "blue", "BunsEaten" **=** "red")) *# Set fill colors*



**Pie Chart**

In [8]:

pie(hotdog\_winners**$**`Dogs eaten`, labels **=** hotdog\_winners**$**Winner, main **=** "Hotdog Contest Wins") *# Set title*



**Create donut chart**

In [9]:

*# Prepare data for donut chart*

donut\_data **<-** hotdog\_winners **%>%**

mutate(fraction **=** `Dogs eaten` **/** sum(`Dogs eaten`))

ggplot(donut\_data, aes(x **=** 2, y **=** fraction, fill **=** Winner)) **+**

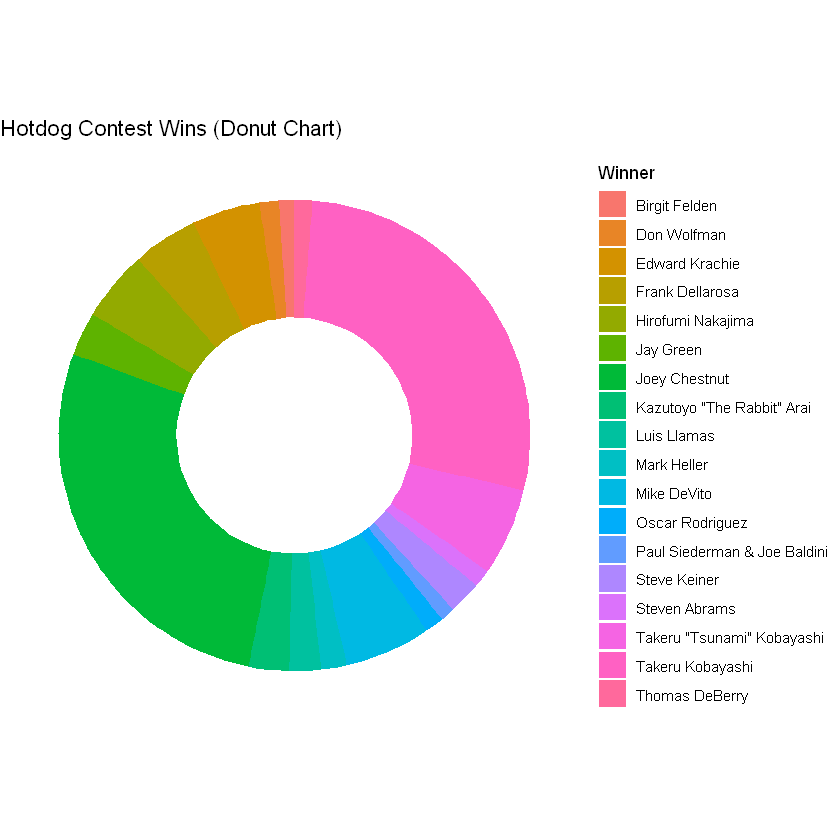
geom\_bar(stat **=** "identity", width **=** 1) **+**

coord\_polar(theta **=** "y") **+**

xlim(0.5, 2.5) **+**

ggtitle('Hotdog Contest Wins (Donut Chart)') **+** *# Set title*

theme\_void() *# Remove background and axes*



**Create line chart**

In [15]:

*# Reshape data from wide to long format*

obama\_approval\_long **<-** pivot\_longer(obama\_approval,

cols **=** **-**Issue,

names\_to **=** "Approval\_Status",

values\_to **=** "Percentage")

*# Print the first few rows of the reshaped data to verify its structure*

head(obama\_approval\_long)

*# Create the line chart*

ggplot(obama\_approval\_long, aes(x **=** Approval\_Status, y **=** Percentage, color **=** Issue, group **=** Issue)) **+**

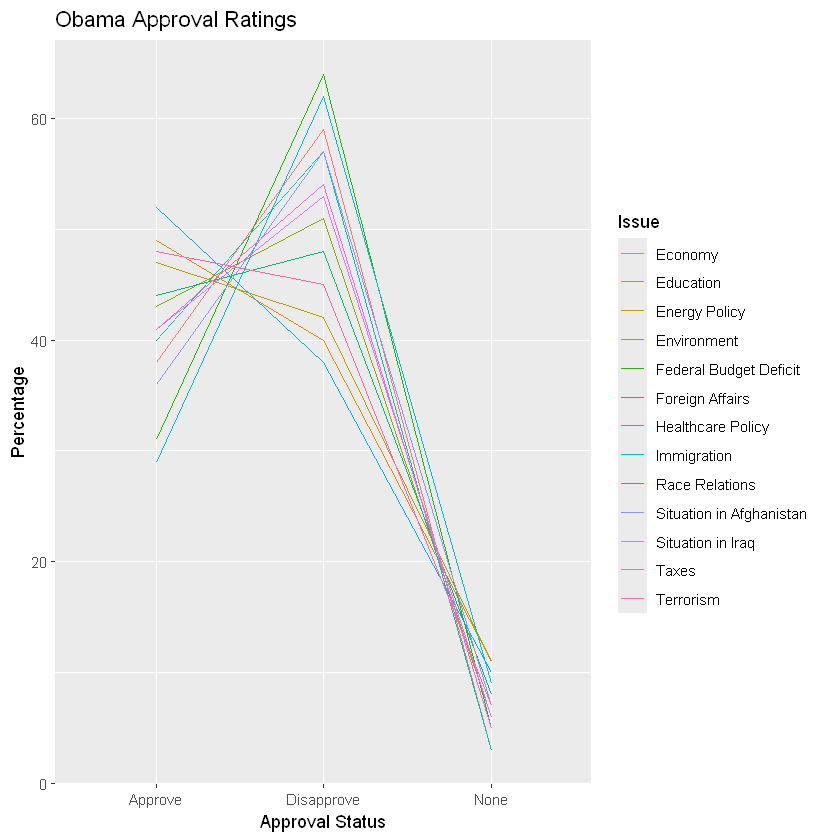
geom\_line() **+**

ggtitle('Obama Approval Ratings') **+**

xlab('Approval Status') **+**

ylab('Percentage')

| A tibble: 6 × 3 | | |
| --- | --- | --- |
| **Issue** | **Approval\_Status** | **Percentage** |
| **<chr>** | **<chr>** | **<dbl>** |
| Race Relations | Approve | 52 |
| Race Relations | Disapprove | 38 |
| Race Relations | None | 10 |
| Education | Approve | 49 |
| Education | Disapprove | 40 |
| Education | None | 11 |



In [ ]:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

### Using Python:

#### Install required libraries

In [1]:

pip install xlrd openpyxl

Requirement already satisfied: xlrd in c:\users\mcken\anaconda3\lib\site-packages (2.0.1)

Requirement already satisfied: openpyxl in c:\users\mcken\anaconda3\lib\site-packages (3.0.10)

Requirement already satisfied: et\_xmlfile in c:\users\mcken\anaconda3\lib\site-packages (from openpyxl) (1.1.0)

Note: you may need to restart the kernel to use updated packages.

#### Import the Data

In [2]:

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

**import** pandas **as** pd

*# Paths to the data files*

obama\_approval\_path **=** "C:/Users/mcken/Downloads/obama-approval-ratings.xls"

hotdog\_places\_path **=** "C:/Users/mcken/Downloads/hotdog-places.xlsm"

hotdog\_winners\_path **=** "C:/Users/mcken/Downloads/hotdog-contest-winners.xlsm"

*# Load the data into pandas DataFrames*

obama\_approval **=** pd**.**read\_excel(obama\_approval\_path, engine**=**'xlrd') *# Load .xls file with xlrd*

hotdog\_places **=** pd**.**read\_excel(hotdog\_places\_path, engine**=**'openpyxl') *# Load .xlsm file with openpyxl*

hotdog\_winners **=** pd**.**read\_excel(hotdog\_winners\_path, engine**=**'openpyxl') *# Load .xlsm file with openpyxl*

*# Print the column names and a few rows of each DataFrame*

print("Obama Approval Ratings Columns:", obama\_approval**.**columns)

print(obama\_approval**.**head())

print("\nHotdog Places Columns:", hotdog\_places**.**columns)

print(hotdog\_places**.**head())

print("\nHotdog Winners Columns:", hotdog\_winners**.**columns)

print(hotdog\_winners**.**head())

Obama Approval Ratings Columns: Index(['Issue', 'Approve', 'Disapprove', 'None'], dtype='object')

Issue Approve Disapprove None

0 Race Relations 52 38 10

1 Education 49 40 11

2 Terrorism 48 45 7

3 Energy Policy 47 42 11

4 Foreign Affairs 44 48 8

Hotdog Places Columns: Index([2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010], dtype='int64')

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

Hotdog Winners Columns: Index(['Year', 'Winner', 'Dogs eaten', 'Country', 'New record'], dtype='object')

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

### Creating the required Charts:

#### Bar Chart

The hotdog\_places DataFrame is melted to make it suitable for plotting.

In [3]:

plt**.**figure(figsize**=**(10, 6)) *# Set the figure size*

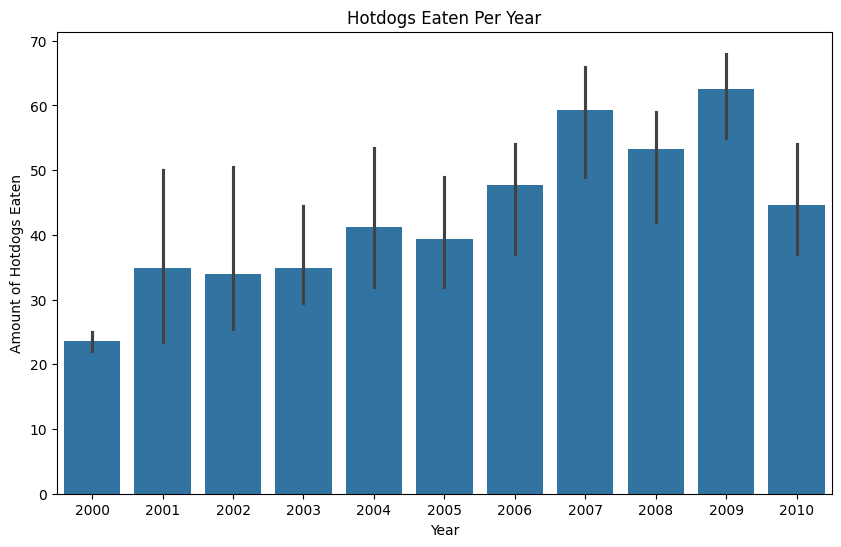
sns**.**barplot(data**=**hotdog\_places**.**melt(), x**=**'variable', y**=**'value') *# Create bar plot using melted DataFrame*

plt**.**title('Hotdogs Eaten Per Year') *# Set title*

plt**.**xlabel('Year') *# Set x-axis label*

plt**.**ylabel('Amount of Hotdogs Eaten') *# Set y-axis label*

plt**.**show() *# Display the plot*



#### Stacked Bar Chart

Dummy data for BunsEaten is generated for the stacked bar chart

In [4]:

hotdog\_places\_stacked **=** hotdog\_places**.**melt(var\_name**=**'Year', value\_name**=**'HotdogsEaten') *# Melt the DataFrame*

hotdog\_places\_stacked['BunsEaten'] **=** hotdog\_places\_stacked['HotdogsEaten'] **\*** 0.8 *# Create dummy 'BunsEaten' data*

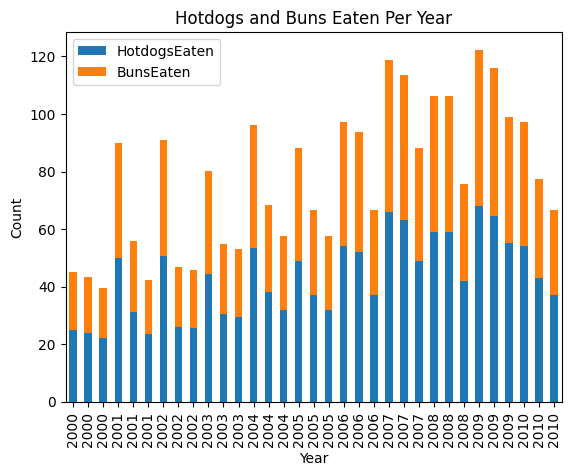
hotdog\_places\_stacked**.**plot(kind**=**'bar', stacked**=True**, x**=**'Year', y**=**['HotdogsEaten', 'BunsEaten']) *# Create stacked bar plot*

plt**.**title('Hotdogs and Buns Eaten Per Year') *# Set title*

plt**.**xlabel('Year') *# Set x-axis label*

plt**.**ylabel('Count') *# Set y-axis label*

plt**.**show() *# Display the plot*



#### Pie Chart

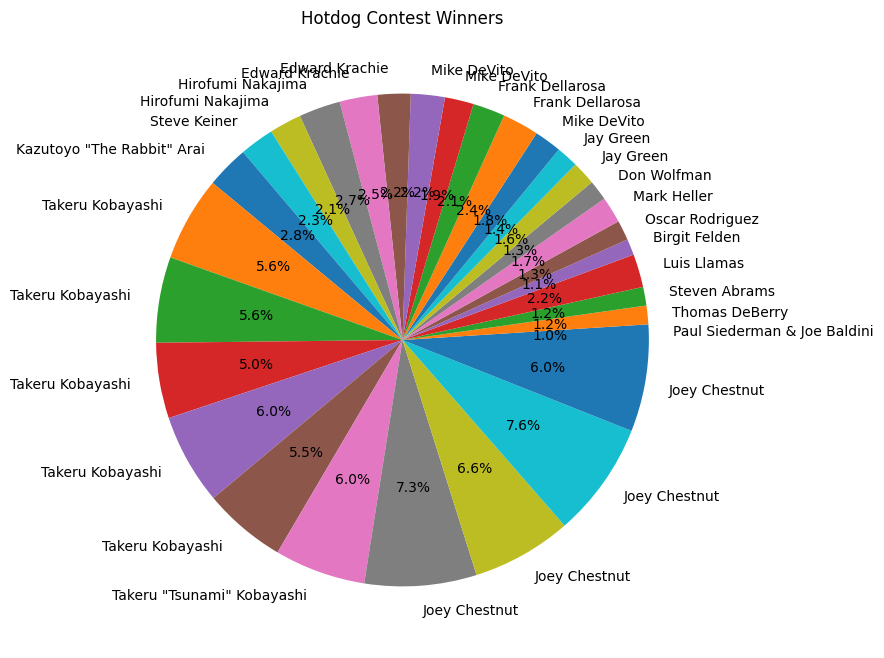
In [5]:

plt**.**figure(figsize**=**(8, 8)) *# Set the figure size*

plt**.**pie(hotdog\_winners['Dogs eaten'], labels**=**hotdog\_winners['Winner'], autopct**=**'%1.1f%%') *# Create pie chart*

plt**.**title('Hotdog Contest Winners') *# Set title*

plt**.**show() *# Display the plot*



#### Donut Chart

In [6]:

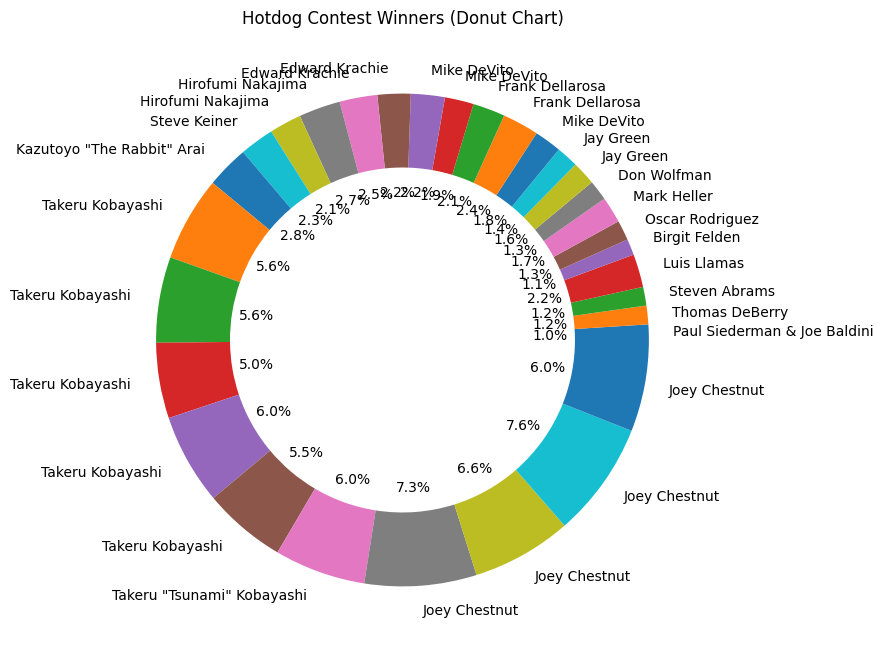
plt**.**figure(figsize**=**(8, 8)) *# Set the figure size*

wedges, texts, autotexts **=** plt**.**pie(hotdog\_winners['Dogs eaten'], labels**=**hotdog\_winners['Winner'], autopct**=**'%1.1f%%') *# Create pie chart*

plt**.**setp(wedges, width**=**0.3) *# Adjust the width to create a donut chart*

plt**.**title('Hotdog Contest Winners (Donut Chart)') *# Set title*

plt**.**show() *# Display the plot*



#### Line Chart

The obama\_approval DataFrame is indexed by the 'Issue' column and used to plot approval ratings.

In [7]:

plt**.**figure(figsize**=**(10, 6)) *# Set the figure size*

obama\_approval**.**set\_index('Issue')[['Approve', 'Disapprove', 'None']]**.**plot(kind**=**'line') *# Create line chart*

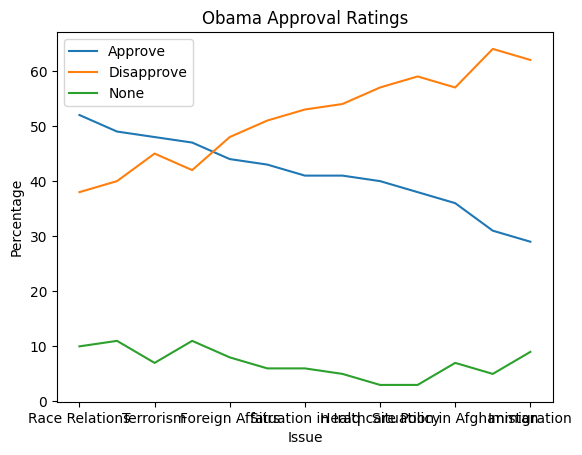
plt**.**title('Obama Approval Ratings') *# Set title*

plt**.**xlabel('Issue') *# Set x-axis label*

plt**.**ylabel('Percentage') *# Set y-axis label*

plt**.**show() *# Display the plot*

<Figure size 1000x600 with 0 Axes>

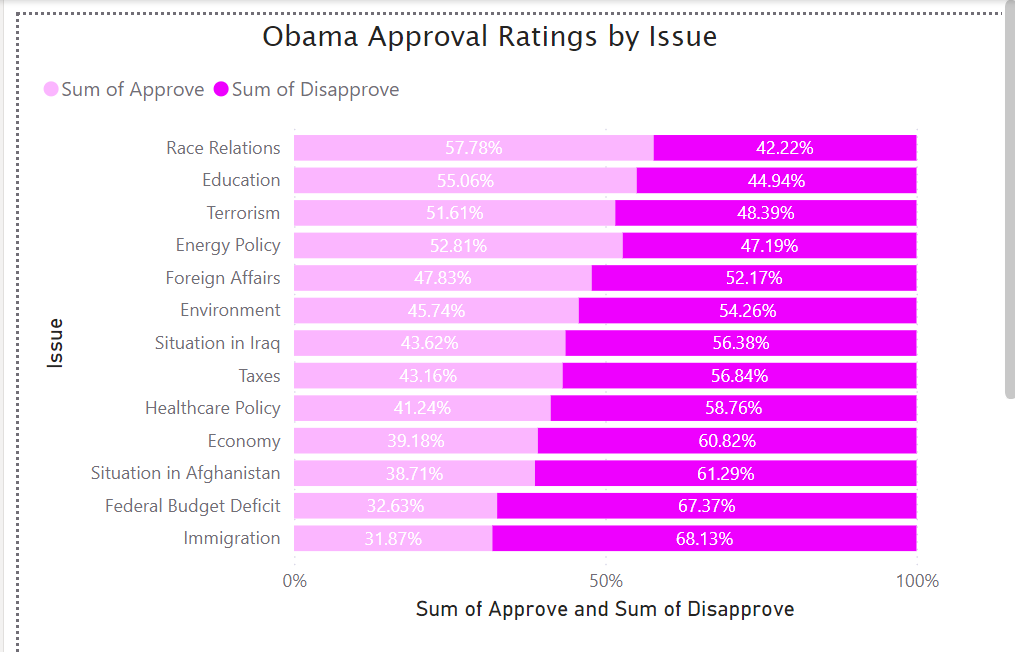


In [ ]:

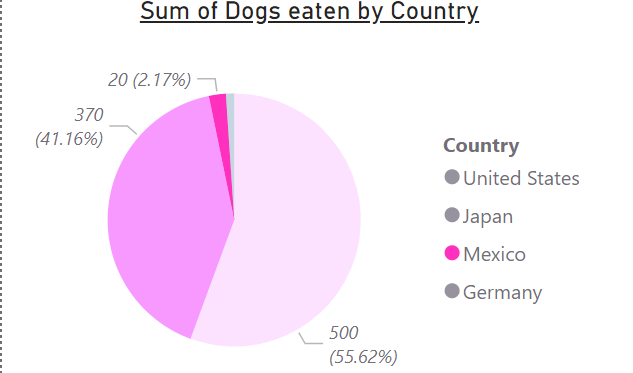
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**\_\_\_Power BI\_\_\_**

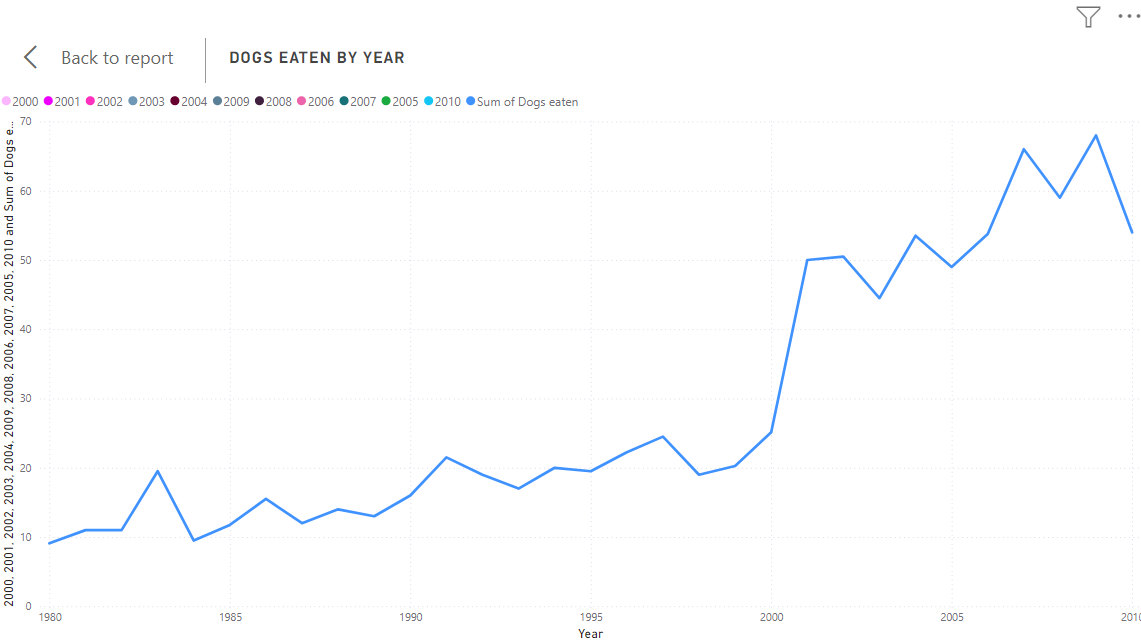
Stacked Bar Chart – Power BI:



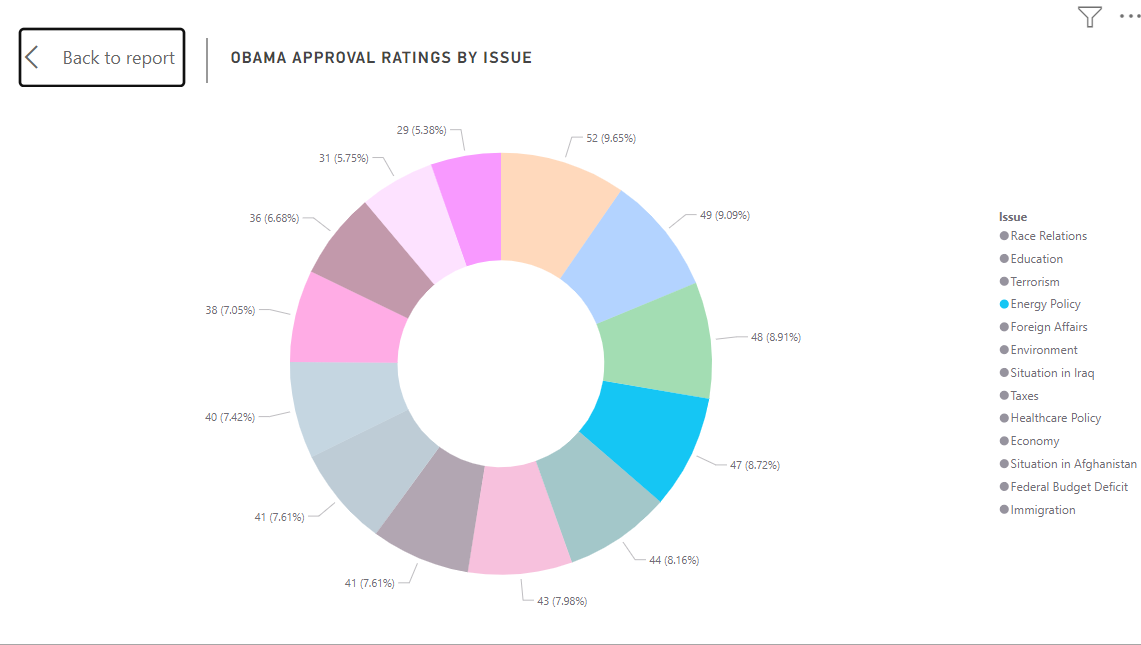
Pie Chart – Power BI:



Line Chart – Power BI:



Donut Chart – Power BI:



Bar Chart – Power BI:

