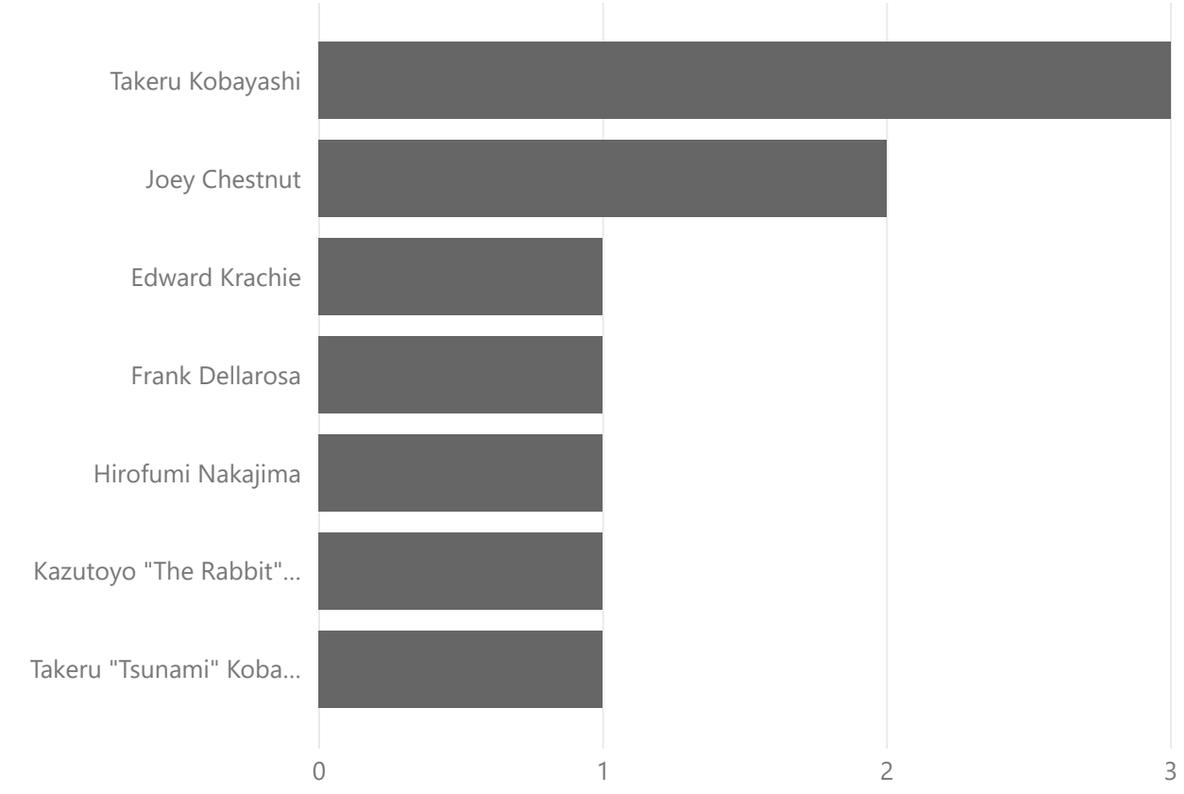
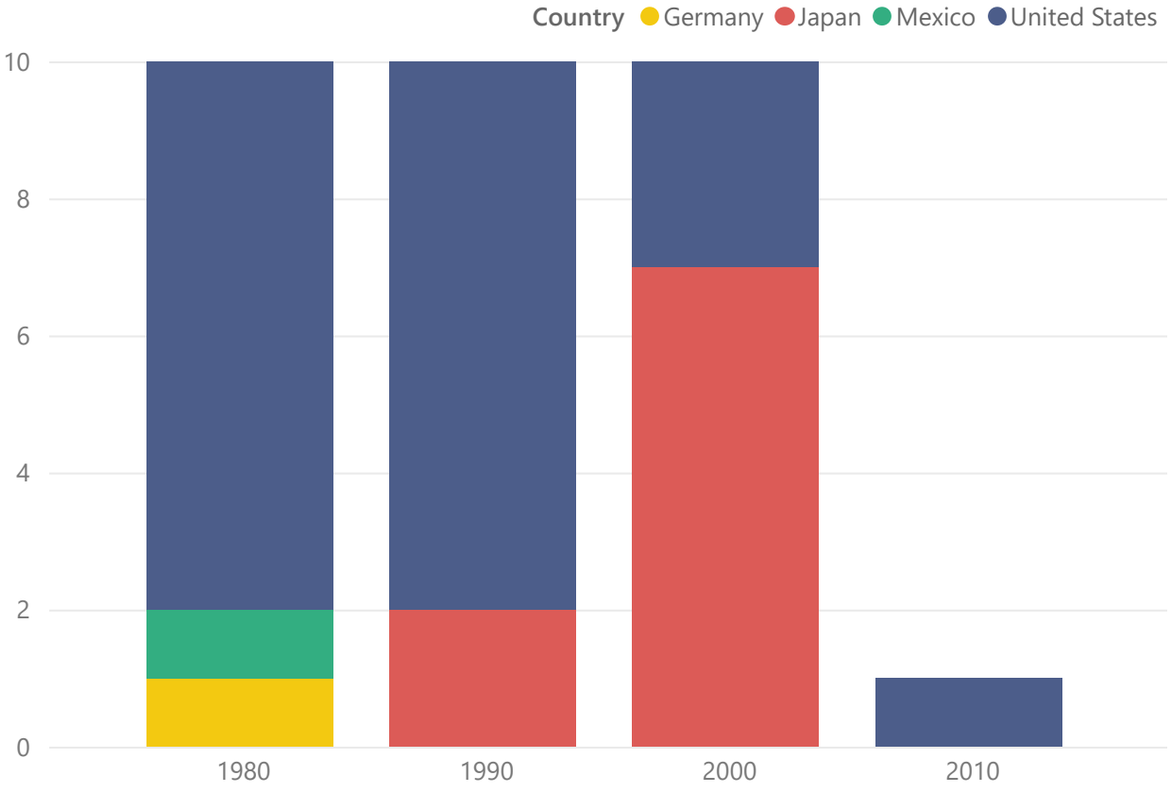
# PowerBI - Bar Chart

Total New Records by Winner



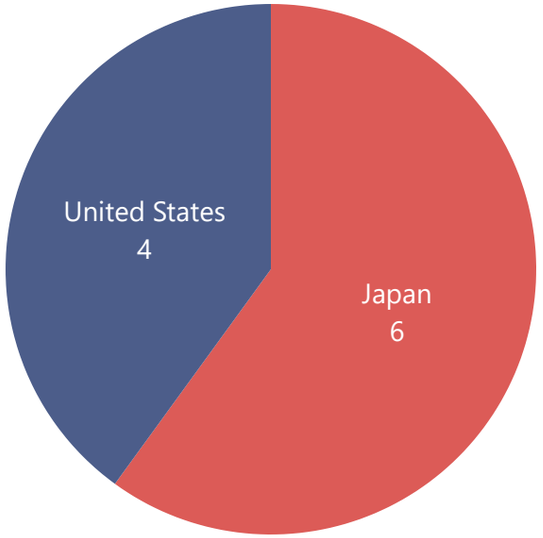

# PowerBI - Stacked Bar Chart

Total Winners by Decade and Country



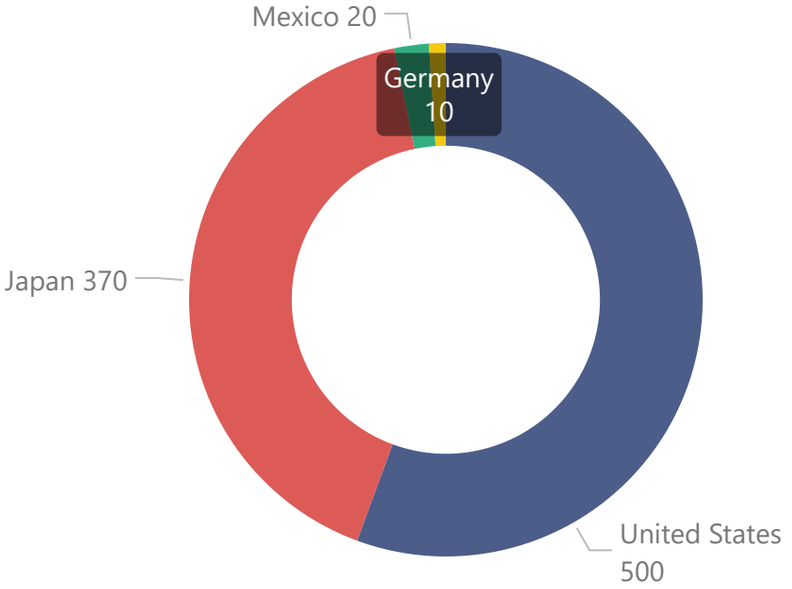
# PowerBI - Pie Chart

Total New Records by Country (1980-2020)




# PowerBI - Donut Chart

Total Hot Dogs eaten by Country (1980-2020)



# Campbell640Week1-2

September 8, 2023

## 1 Assignment 1.2

### 1.1 Import File

[1]:
```
# Load libraries
import pandas as pd
import numpy as np
import math
# import data visualization libraries
import seaborn as sns
import matplotlib.pyplot as plt
```

[2]:
```
# Read csv file
winners = pd.read_excel('ex1-2/hotdog-contest-winners.xlsm')
```

[5]:
```
winners.shape
```

[5]:
```
(31, 5)
```

[6]:
```
winners.head()
```

[6]:

| | 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 |

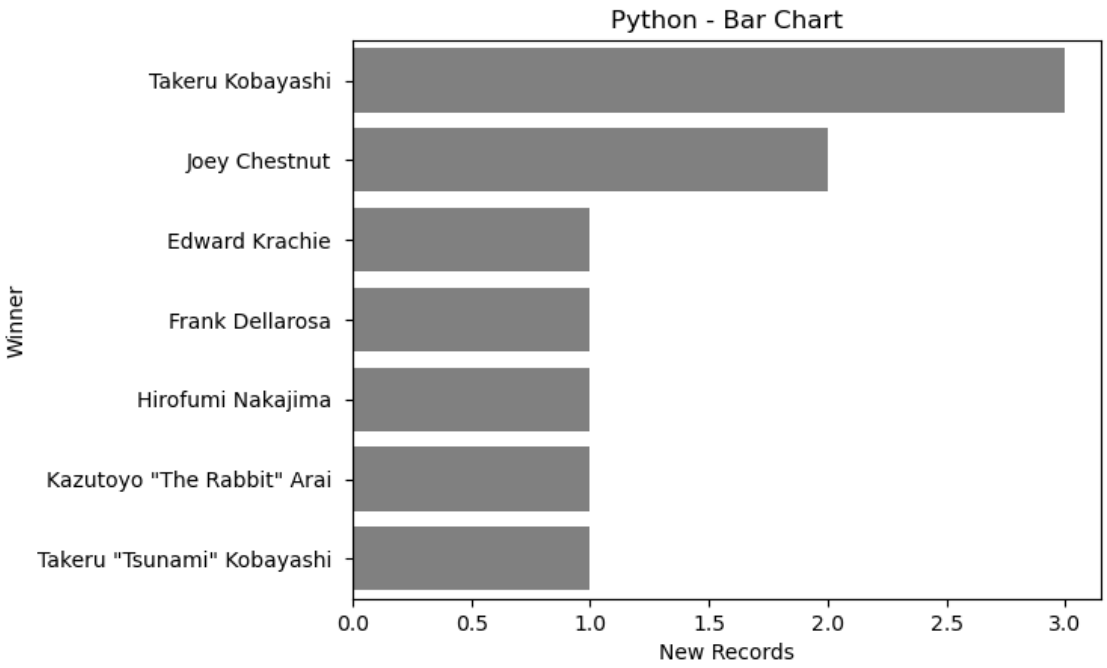
### 1.2 Graphs

#### 1.2.1 Python - Bar Chart

[3]:
```
# New record count by winner name
wnr_df = winners.groupby('Winner', as_index=False)['New record'].sum().
 ↪sort_values('New record', ascending=False)
```

[4]:
```
# Remove 0 values
wnr_df = wnr_df[wnr_df['New record'] != 0]
```

[5]:
```
# Create bar chart
sns.barplot(data=wnr_df, x="New record", y="Winner", orient = 'h', ci=None,␣
 ↪color='gray')
plt.title('Python - Bar Chart')
plt.xlabel('New Records')
plt.ylabel('Winner')
plt.show()
```




### 1.2.2 Python - Stacked Bar Chart

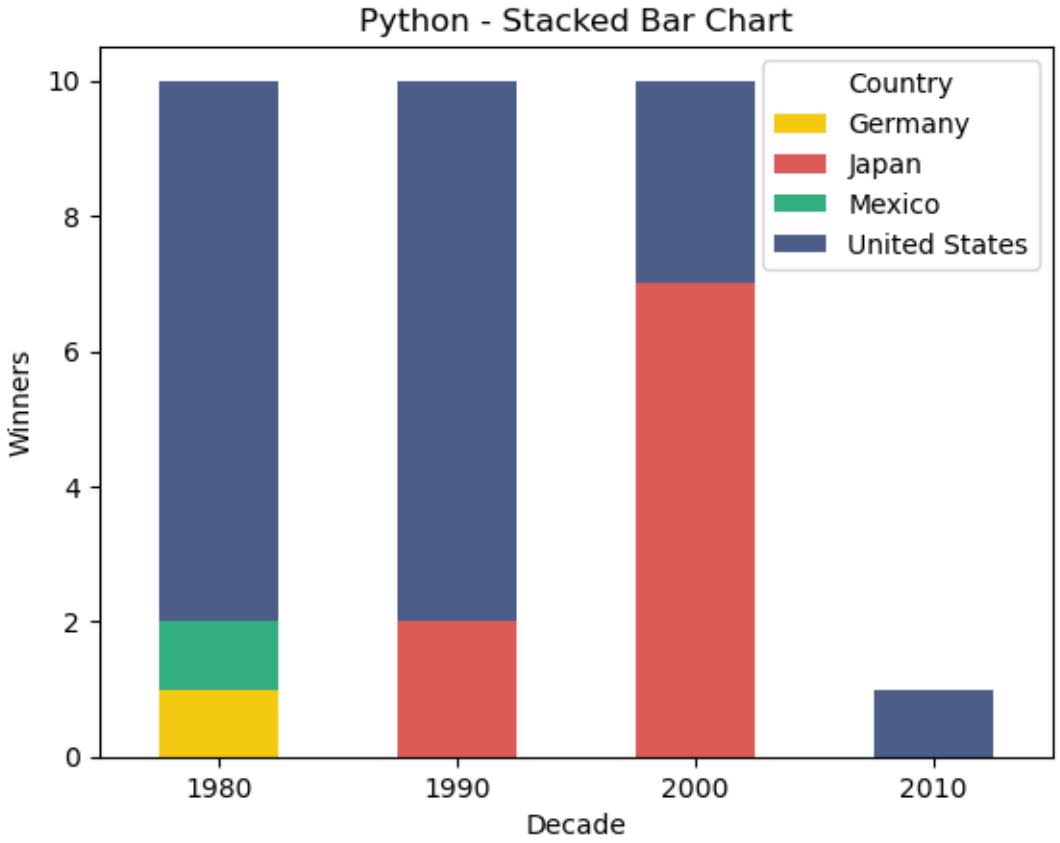
[6]:
```
# Find decade for year
winners['decade'] = np.floor(winners['Year']/10)*10
# Convert to int
winners['decade'] = winners['decade'].astype(int)
```

[7]:
```
# New record count by winner name
cdw_df = winners.groupby(['decade','Country'])['Winner'].count().unstack().
 ↪fillna(0)
```

[8]:
```
color=["#F3C911","#DC5B57","#33AE81","#4C5D8A"]
```

[9]:
```
cdw_df.plot(kind='bar', stacked=True,color=color)
plt.xticks(rotation=0, ha='center')
plt.title('Python - Stacked Bar Chart')
plt.xlabel('Decade')
```

```
plt.ylabel('Winners')
plt.show()
```




### 1.2.3 Python - Pie Chart

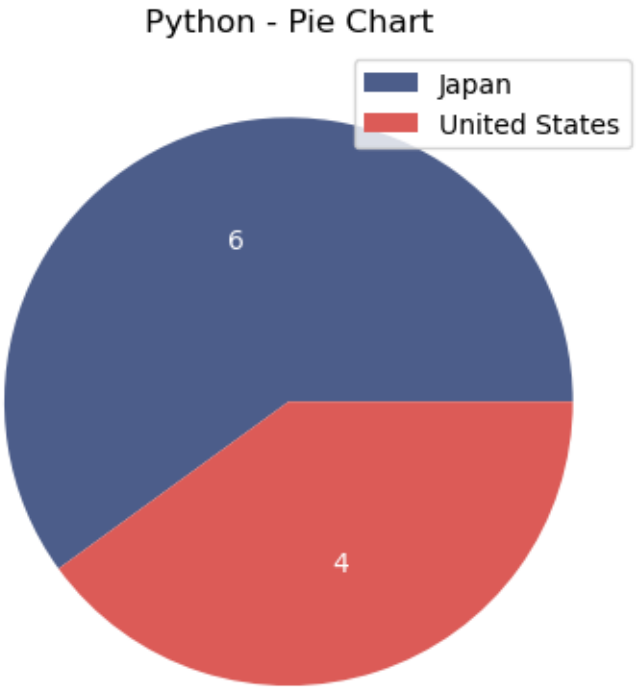
[10]:
```
# New record count by winner name
cnr_df = winners.groupby('Country')['New record'].sum()
```

[11]:
```
# Remove 0 values
cnr_df = cnr_df[cnr_df.iloc[0:4] != 0]
```

[12]:
```
def make_autopct(values):
    def my_autopct(pct):
        total = sum(values)
        val = int(round(pct*total/100.0))
        return '{v:d}'.format(v=val)
    return my_autopct
```

[13]:
```
color=["#4C5D8A","#DC5B57"]
```

[14]:
```
plt.pie(x=cnr_df, autopct=make_autopct(cnr_df),colors=color,textprops={'color':
  ↪"white"})
plt.title('Python - Pie Chart')
plt.legend(labels=cnr_df.index, loc="best")
plt.show()
```

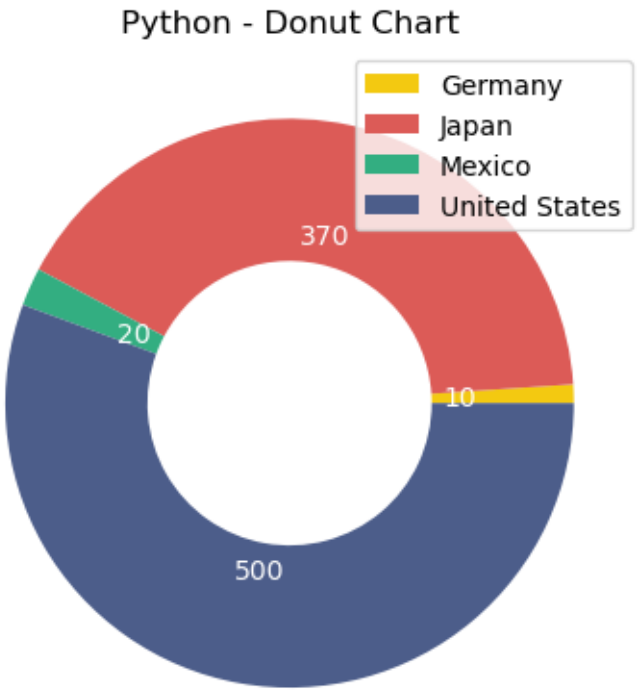



### 1.2.4 Python - Donut Chart

[15]:
```
# New record count by winner name
cde_df = winners.groupby('Country')['Dogs eaten'].sum()
```

[16]:
```
color=["#F3C911","#DC5B57","#33AE81","#4C5D8A"]
```

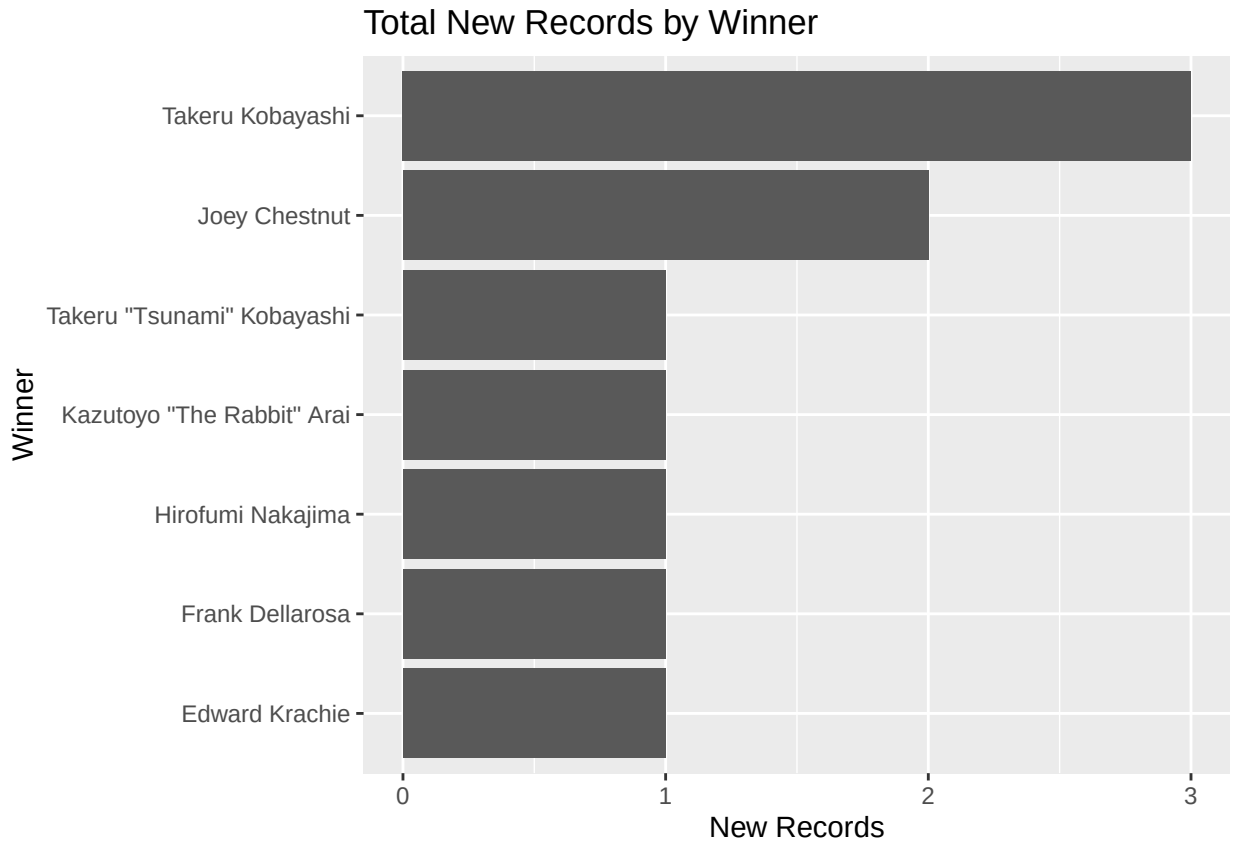
[17]:
```
plt.pie(cde_df, wedgeprops={'width': 0.5},␣
  ↪autopct=make_autopct(cde_df),colors=color,textprops={'color':"white"})
plt.title('Python - Donut Chart')
plt.legend(labels=cde_df.index, loc="best")
plt.show()
```


Python - Donut Chart
Germany
Japan
Mexico
United States
370
20
10
500

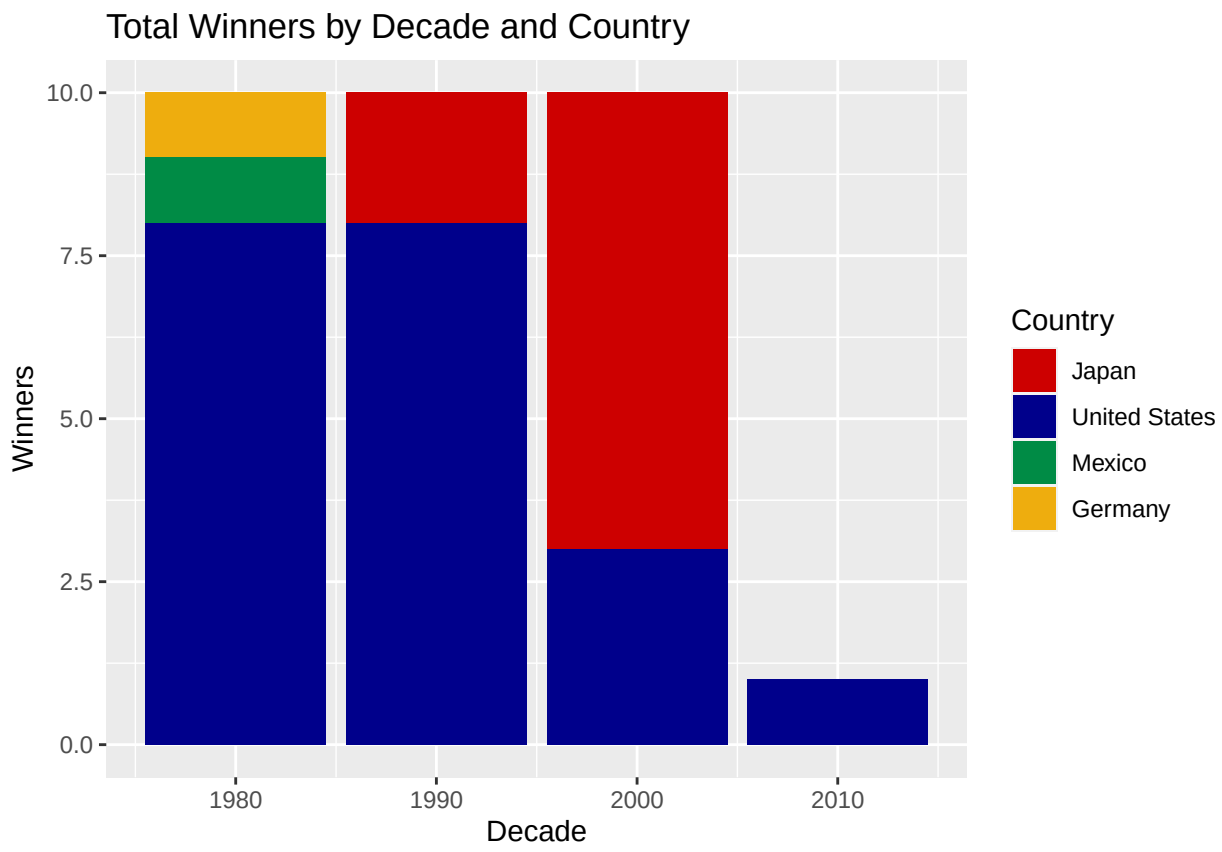
# Campbell640Week1

2023-09-03

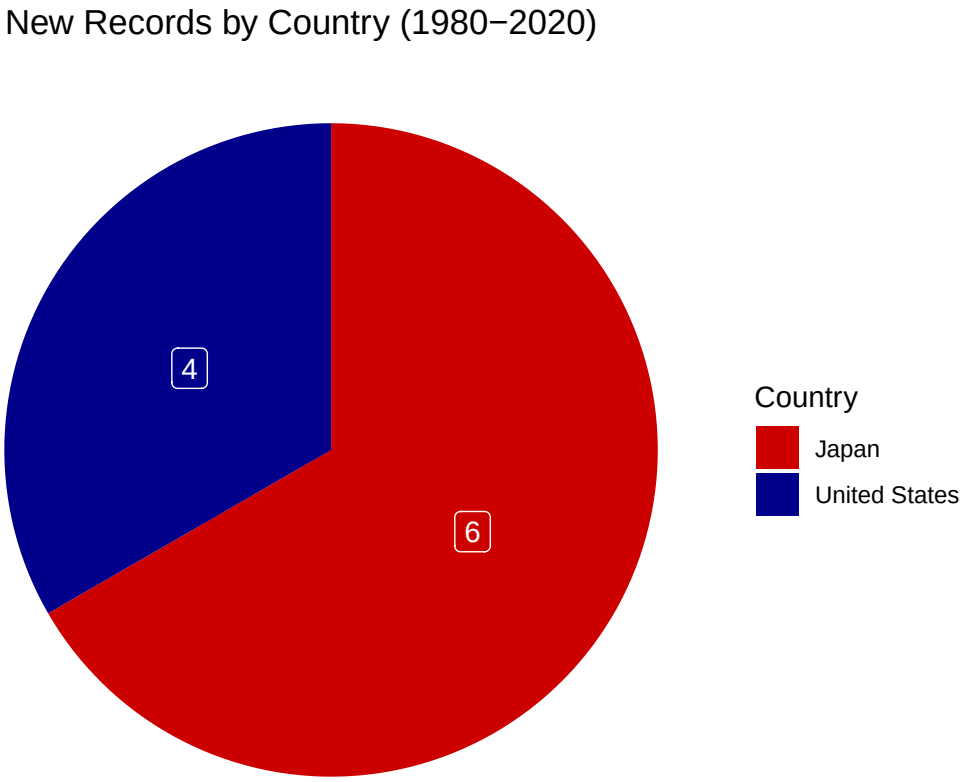
## R - Bar Chart



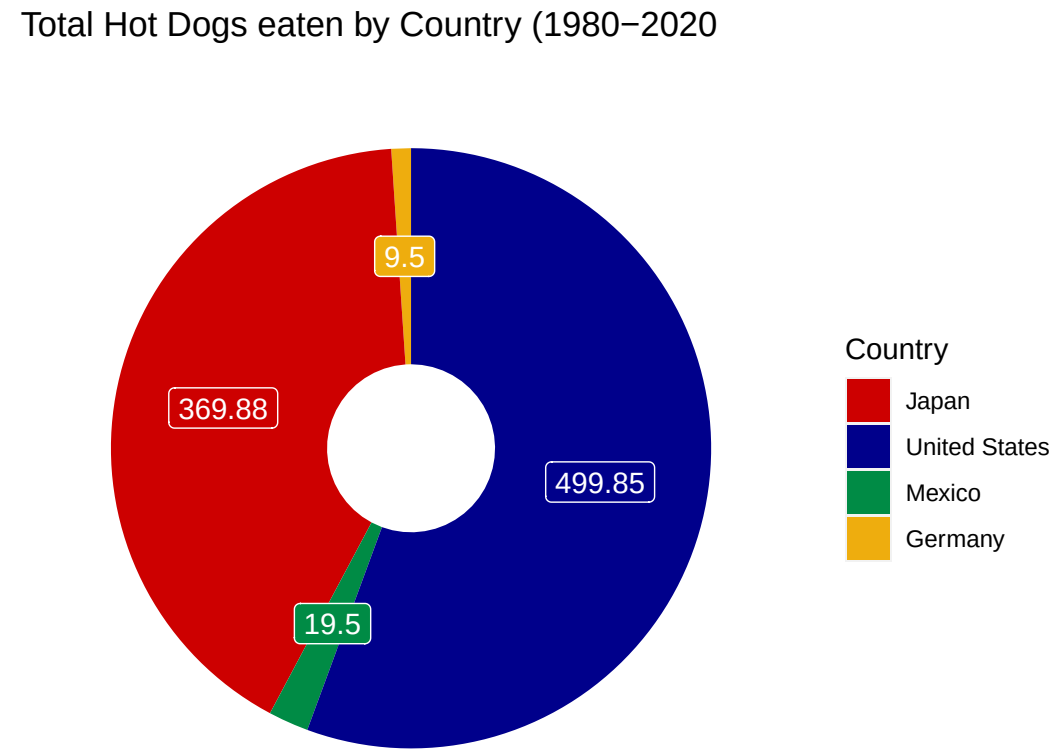
## R - Stacked Bar Chart



## R - Pie Chart


Total New Records by Country (1980−2020)
4
6
Country
Japan
United States

## R - Donut Chart




## Code Repository

```
## Set the working directory to the root of your DSC 640 directory
setwd("C:/Users/jcamp/Documents/DSC640/Assignments/data sources")

# Load libraries
library(readxl)
library(ggplot2)
library(dplyr)
library(tidyverse)

# Load file
contest_winners_df <- read_excel("hotdog-contest-winners.xlsm")

# Adjust new record column name
contest_winners_df <-
  rename(contest_winners_df, new_record = 'New record',
         dogs_eaten = 'Dogs eaten')
# Summarize new records per winner
nr_df = contest_winners_df %>%
  group_by(Winner) %>%
  summarise(count = sum(new_record))

# Remove names with 0 new records
```

```r
nr_df = nr_df[apply(nr_df!=0, 1, all),]

# Create bar chart
ggplot(nr_df, aes(x=count, y = reorder(Winner, count))) +
  geom_col() +
  ggtitle("Total New Records by Winner") +
  xlab("New Records") +
  ylab("Winner")
# Summarize number of winners by country
wc_df = contest_winners_df %>%
  mutate(decade = floor(Year/10)*10) %>%
  group_by(Country, decade) %>%
  tally()

# Create stacked bar chart
ggplot(wc_df, aes(x=decade, y=n, fill=Country)) +
  geom_col() +
  scale_fill_manual(values = c("Japan" = "red3",
                               "United States" = "darkblue",
                               "Mexico" = "springgreen4",
                               "Germany" = "darkgoldenrod2")) +
  ggtitle("Total Winners by Decade and Country") +
  xlab("Decade") +
  ylab("Winners")
# Summarize number of winners by country
cnr_df = contest_winners_df %>%
  group_by(Country) %>%
  summarise(count = sum(new_record))

# Remove names with 0 new records
cnr_df = cnr_df[apply(cnr_df!=0, 1, all),]

# Create pie chart
ggplot(cnr_df, aes(x = "", y = factor(count), fill = Country)) +
  geom_col() +
  coord_polar(theta = "y")  +
  guides(fill = guide_legend(title = "Country")) +
  scale_fill_manual(values = c("red3","darkblue")) +
  geom_label(aes(label = count),
             position = position_stack(vjust = 0.5),
             color = "white",
             show.legend = FALSE) +
  theme_void() +
  ggtitle("Total New Records by Country (1980-2020)")
# Create donut chart

# Summarize number of hot dogs eaten by country
chd_df = contest_winners_df %>%
  group_by(Country) %>%
  summarise(count = sum(dogs_eaten))

# Big hole
hsize <- 1
```

```r
chd_df <- chd_df %>%
  mutate(x = 1)

ggplot(chd_df, aes(x=hsize, y = count, fill = Country)) +
  geom_col() +
  geom_label(aes(label = count),
             position = position_stack(vjust = 0.5),
             color="white",
             show.legend = FALSE) +
  coord_polar(theta = "y") +
  xlim(c(0.2, hsize + 0.5))+
  scale_fill_manual(values = c("Japan" = "red3",
                               "United States" = "darkblue",
                               "Mexico" = "springgreen4",
                               "Germany" = "darkgoldenrod2")) +
  theme(panel.background = element_rect(fill = "white"),
        panel.grid = element_blank(),
        axis.title = element_blank(),
        axis.ticks = element_blank(),
        axis.text = element_blank())+
  ggtitle("Total Hot Dogs eaten by Country (1980-2020")
```