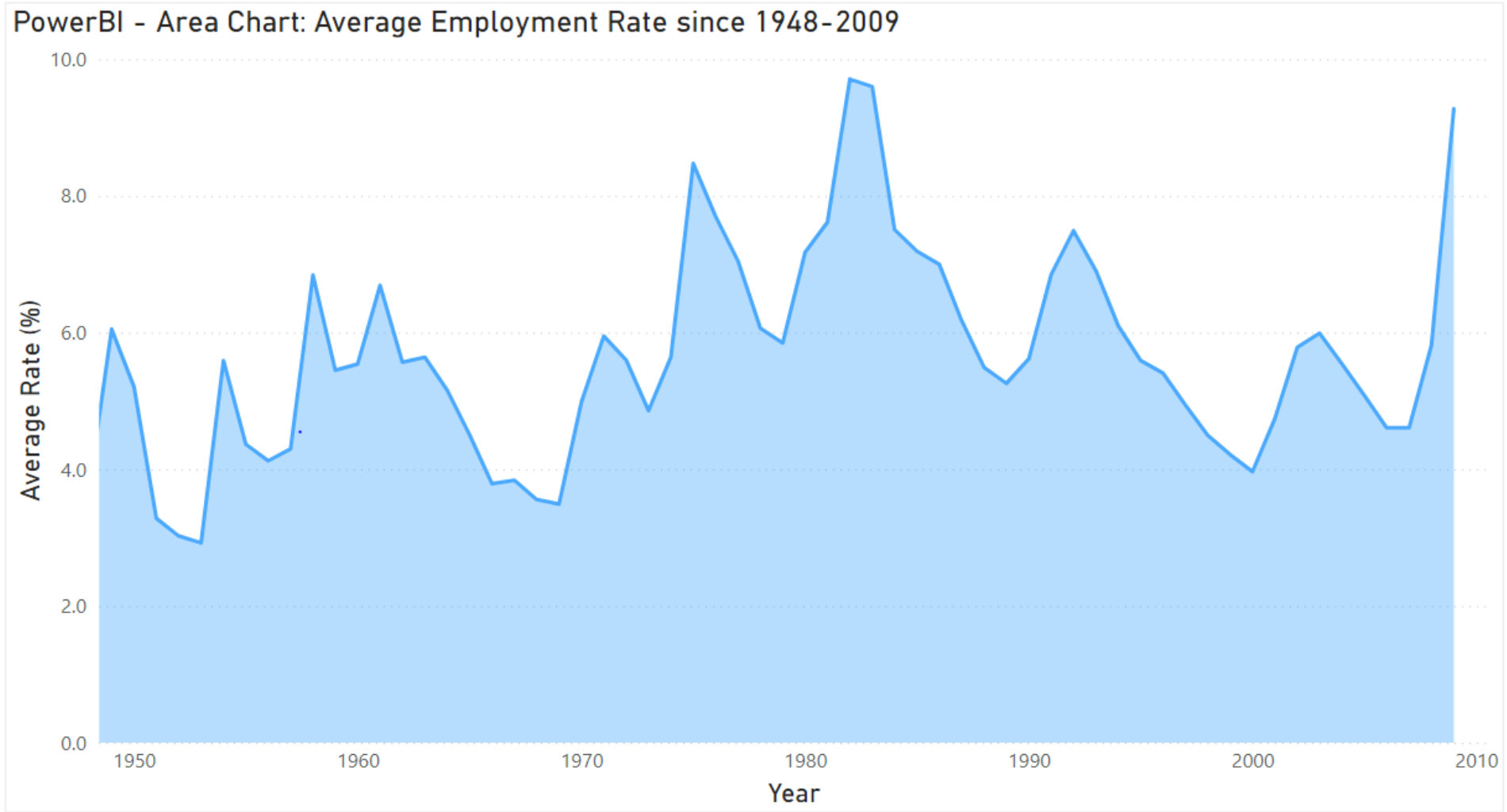
Assignment 3.2 Charts

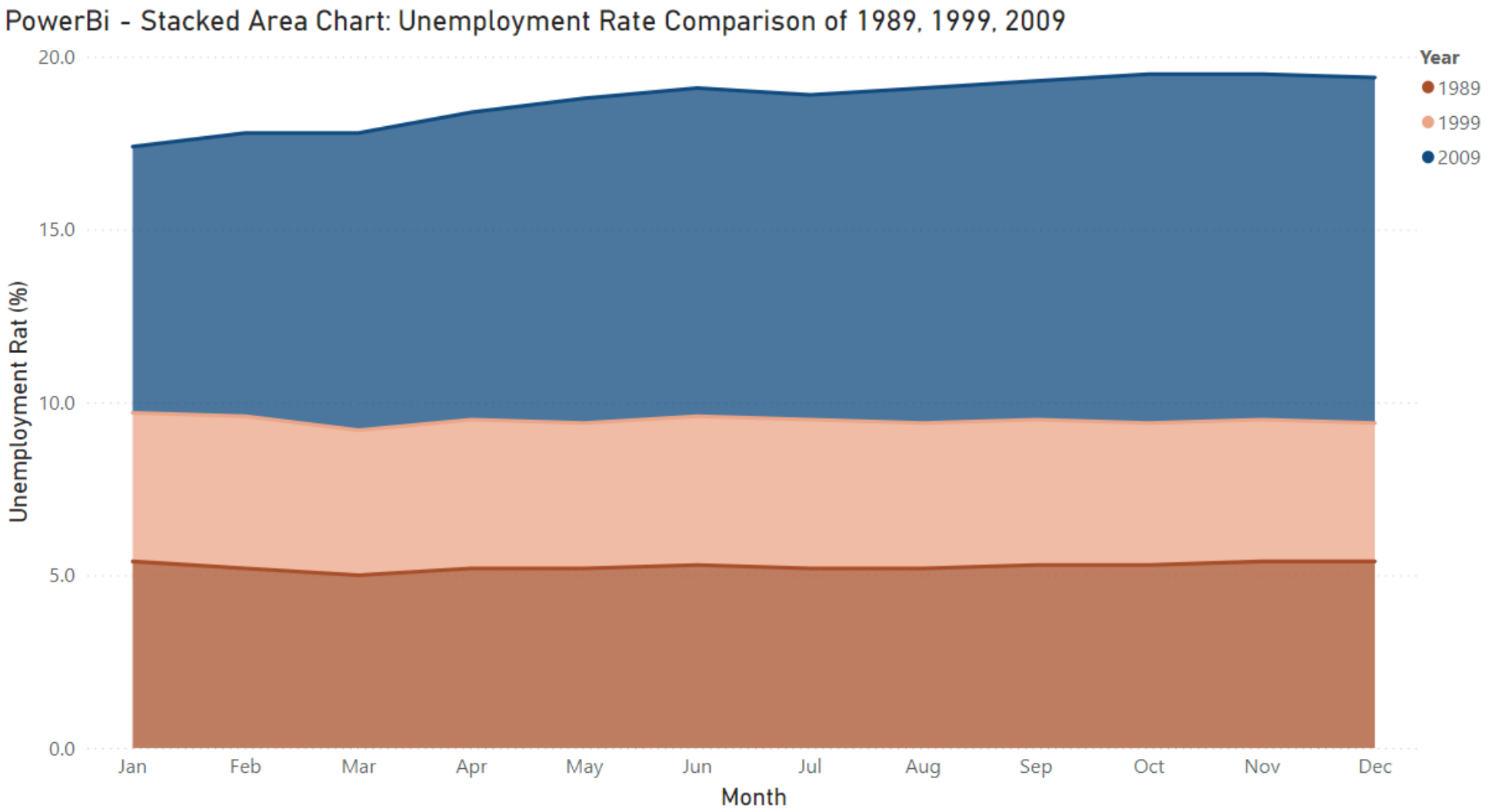
DSC640

Taniya Adhikari

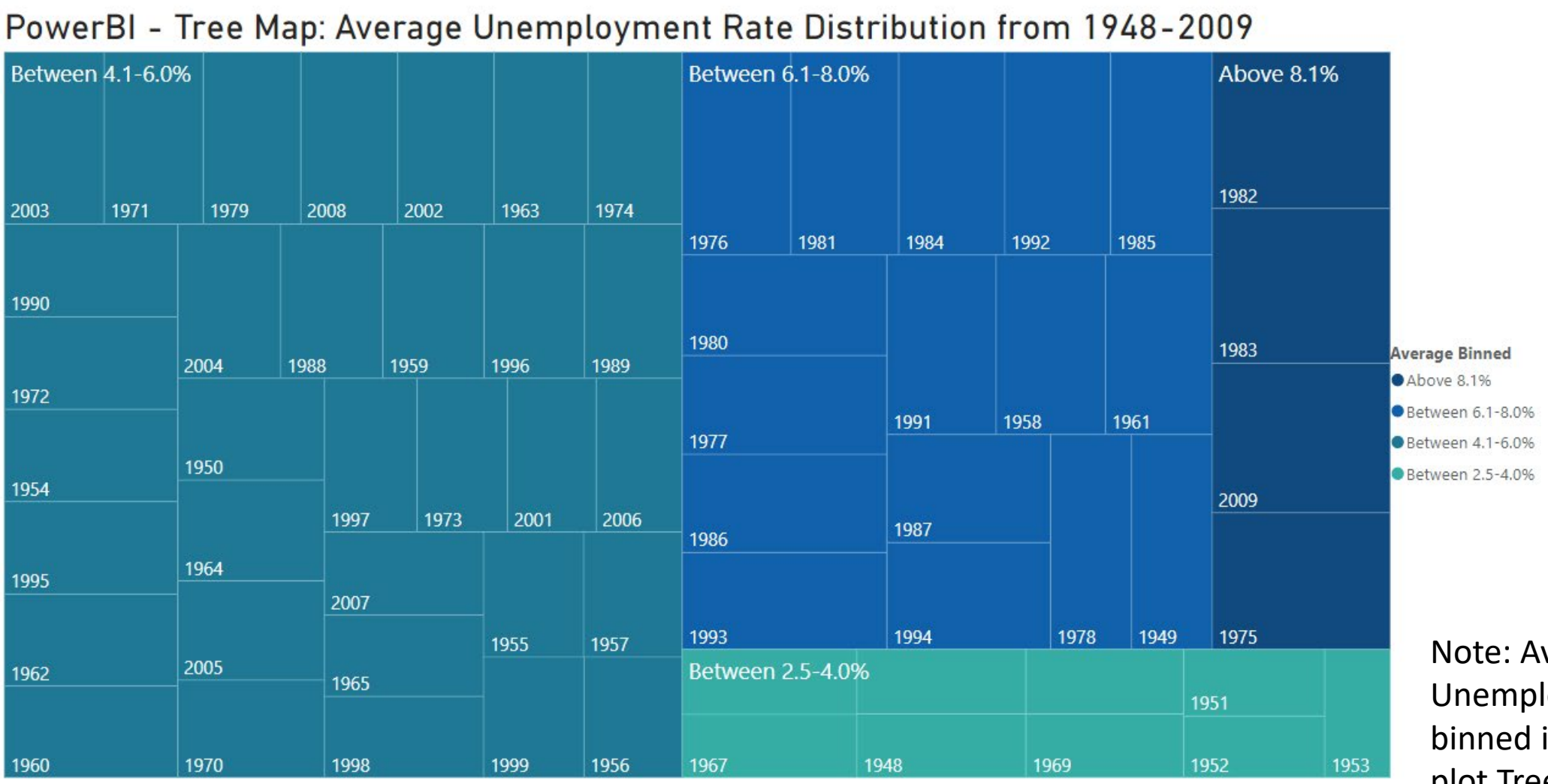
PowerBI – Area Chart



PowerBI – Stacked Area Chart

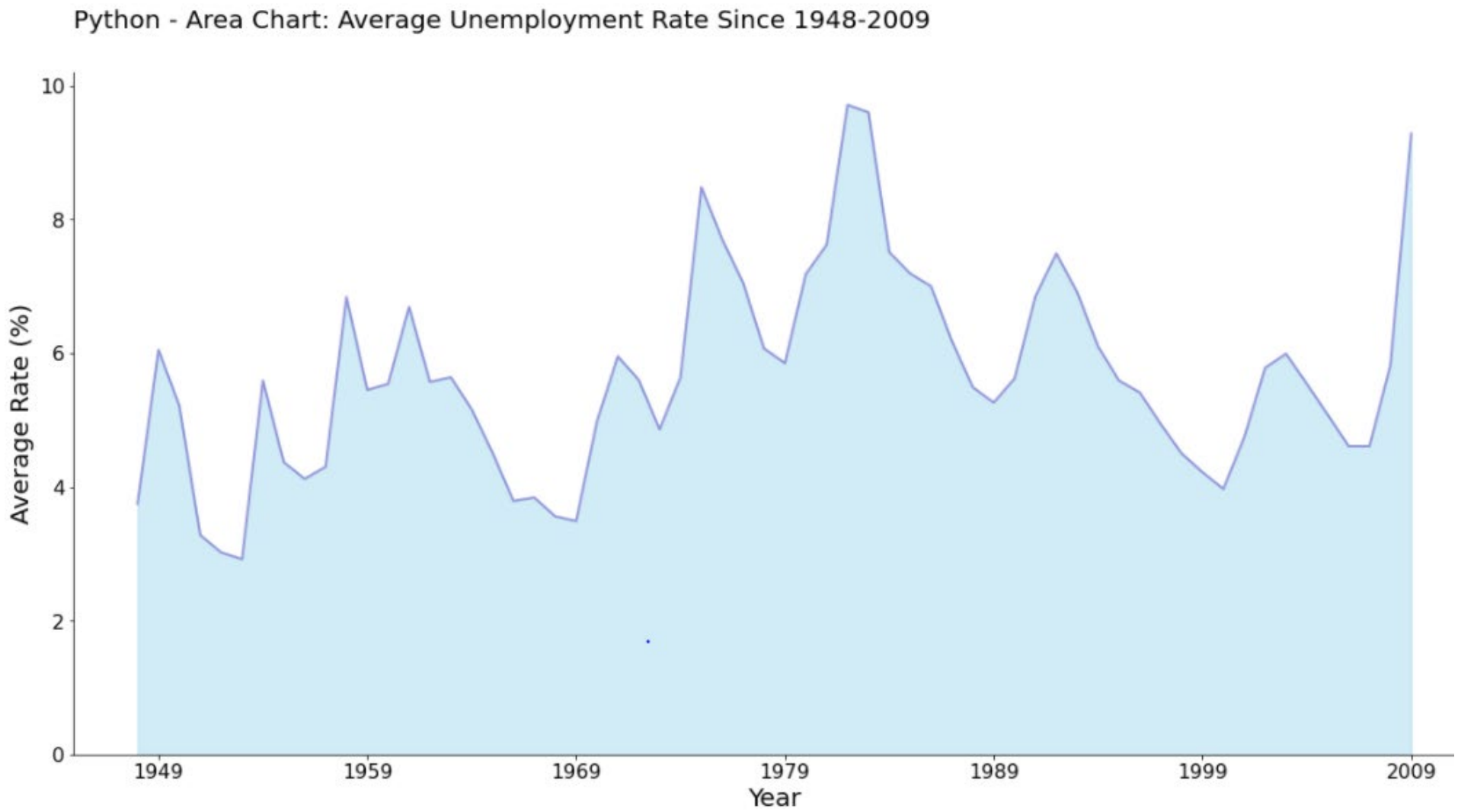


PowerBI – Tree Map

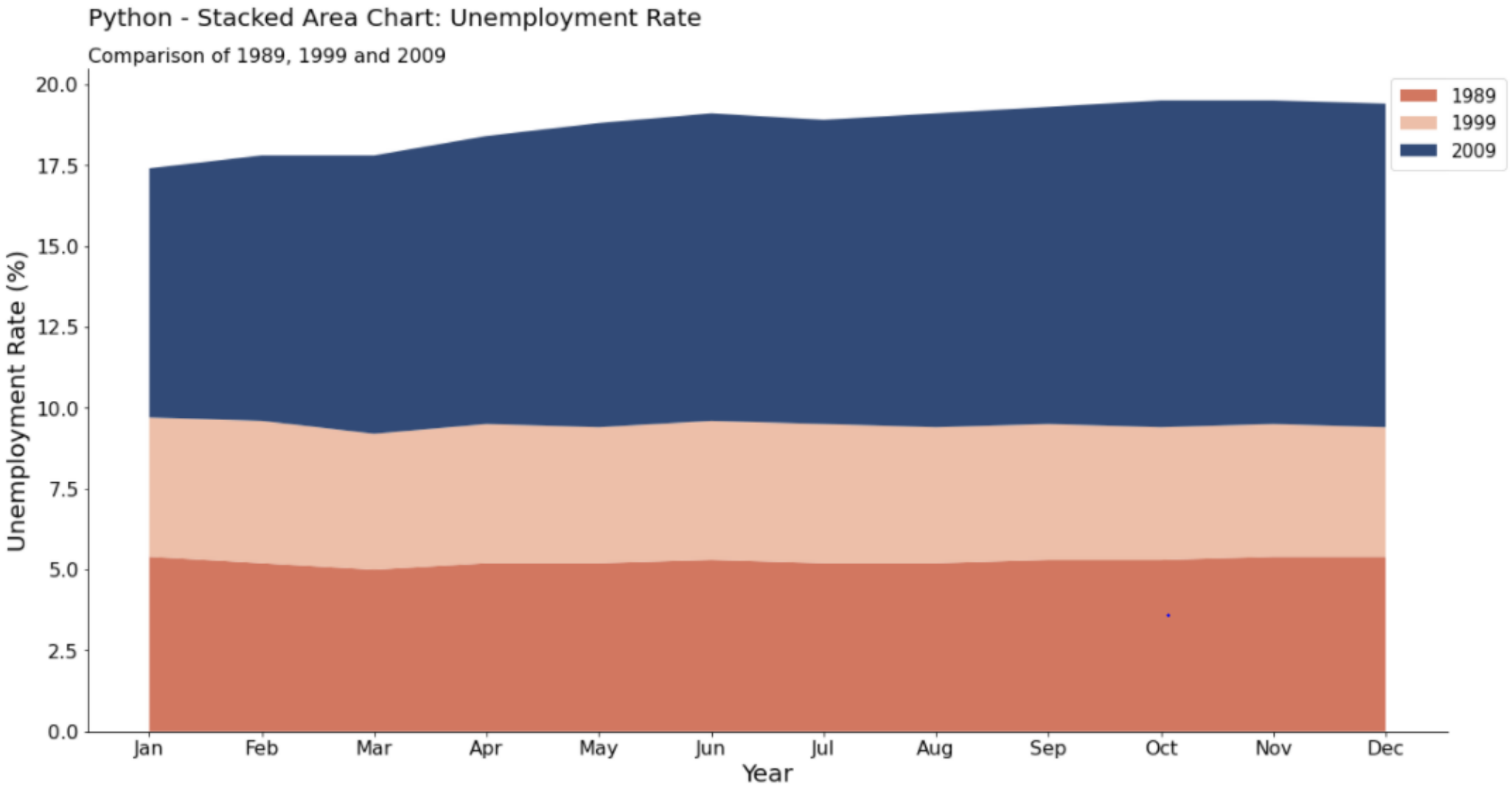


Note: Average Unemployment Rate is binned into categories to plot Tree Maps

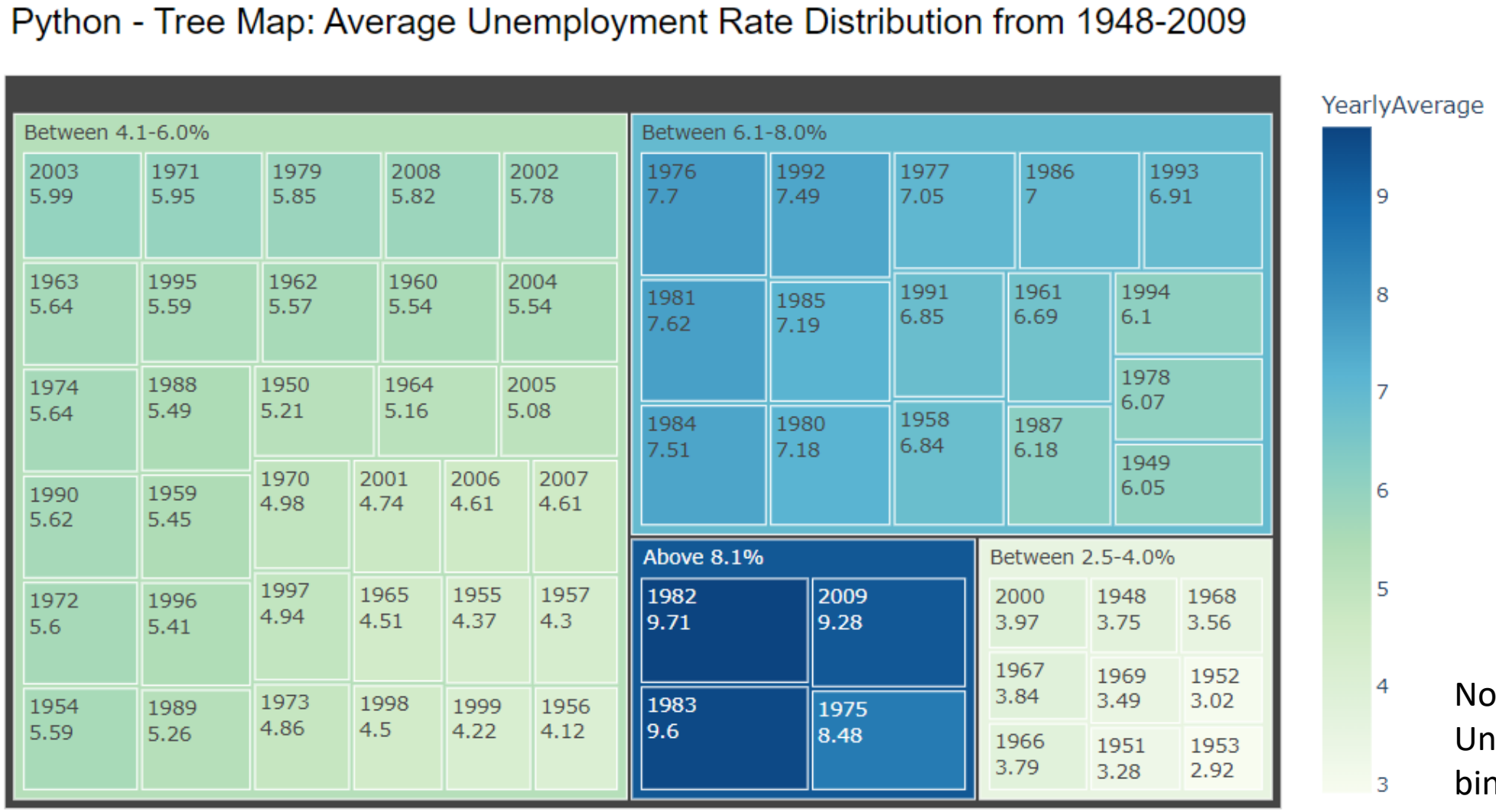
Python – Area Chart



Python – Stacked Area Chart

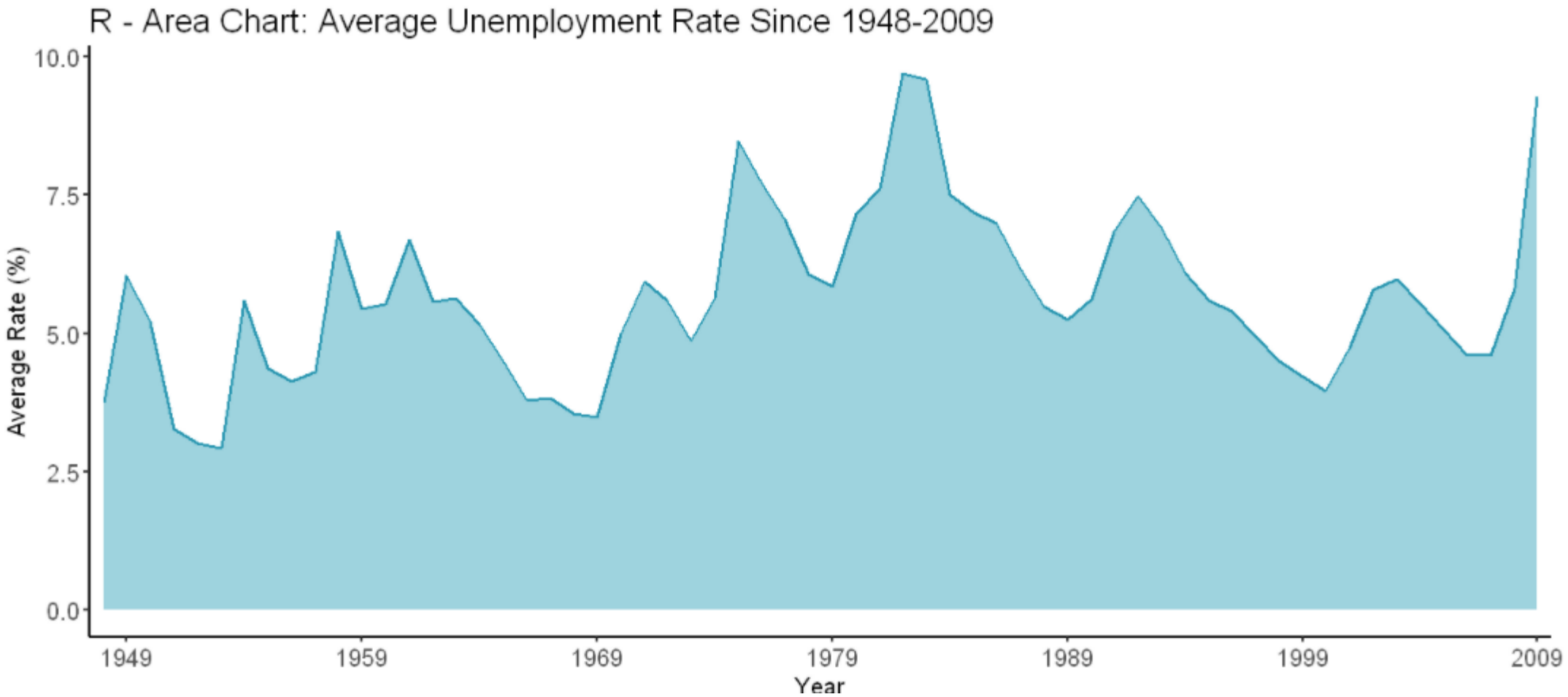


Python – Tree Map

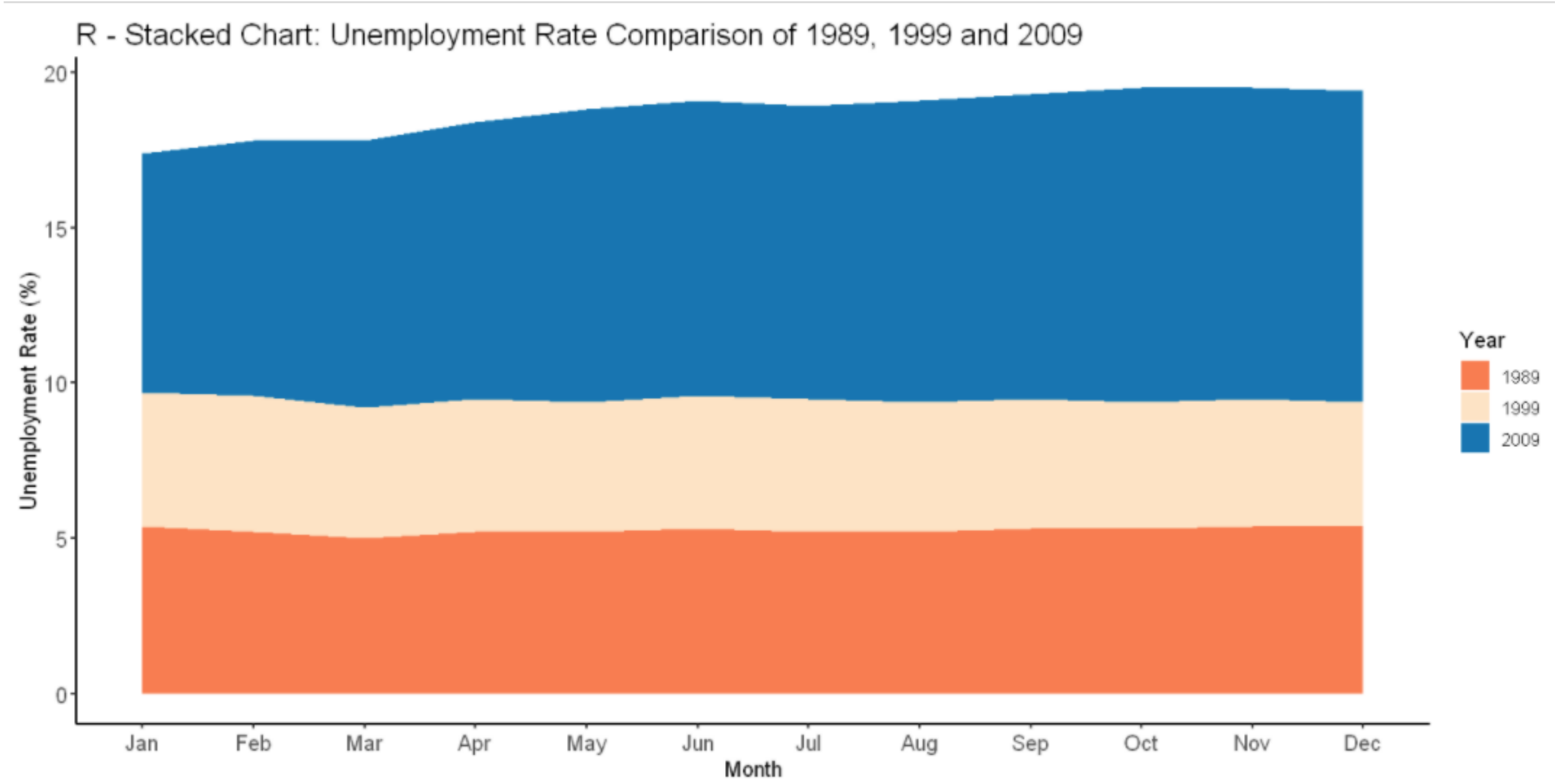


Note: Average Unemployment Rate is binned into categories to plot Tree Maps

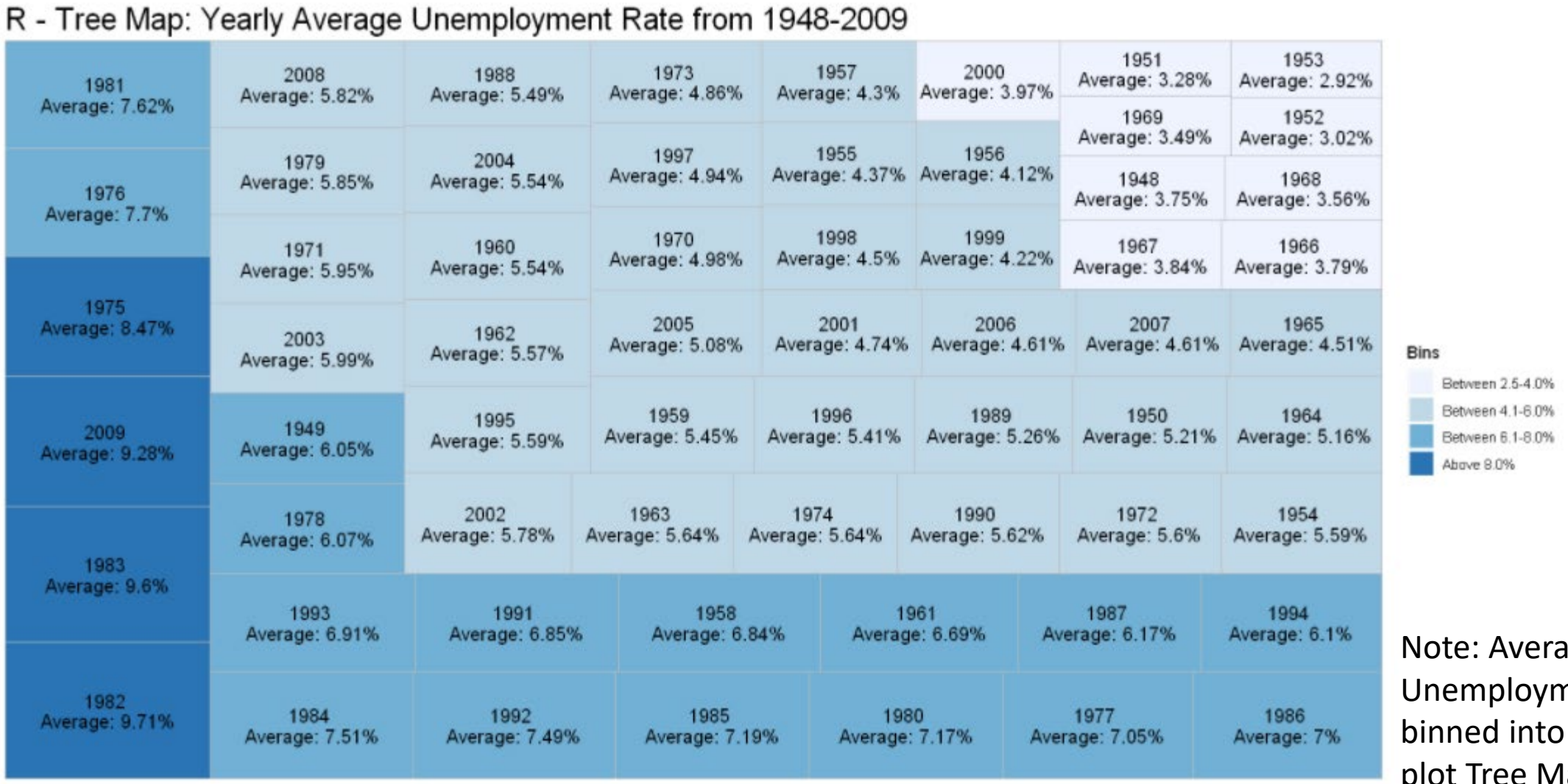
R – Area Chart


R - Area Chart: Average Unemployment Rate Since 1948-2009
Average Rate (%)
10.0
7.5
5.0
2.5
0.0
1949
1959
1969
1979
1989
1999
2009
Year

R – Stacked Area Chart



R – Tree Map



Note: Average Unemployment Rate is binned into categories to plot Tree Maps

Supplemental Files

- Python Code
- R Code
- PowerBI Screenshot

Python Script

Assignment 3.2: Tree Maps, Area Chart and Stacked Area Chart

DSC640

Taniya Adhikari

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
from numerize import numerize
import matplotlib.ticker as ticker
from matplotlib.ticker import FuncFormatter
import plotly.express as px
```

In [2]:
```
df = pd.read_csv('unemployement-rate-1948-2010.csv')
```

In [3]:
```
df['Period'].replace({"M01": "Jan", "M02": "Feb","M03": "Mar","M04": "Apr","M05": "May","M06": "Jun",
                      "M07": "Jul", "M08": "Aug","M09": "Sep","M10": "Oct","M11": "Nov","M12": "Dec"}, inplace=True)
```

In [4]:
```
df.describe()
```

Out[4]:

	Year	Value
count	746.000000	746.000000
mean	1978.584450	5.666488
std	17.957638	1.567909
min	1948.000000	2.500000
25%	1963.000000	4.525000
50%	1979.000000	5.500000
75%	1994.000000	6.600000
max	2010.000000	10.800000

In [5]:
```
df[df['Year'] ==2010]
```

Out[5]:

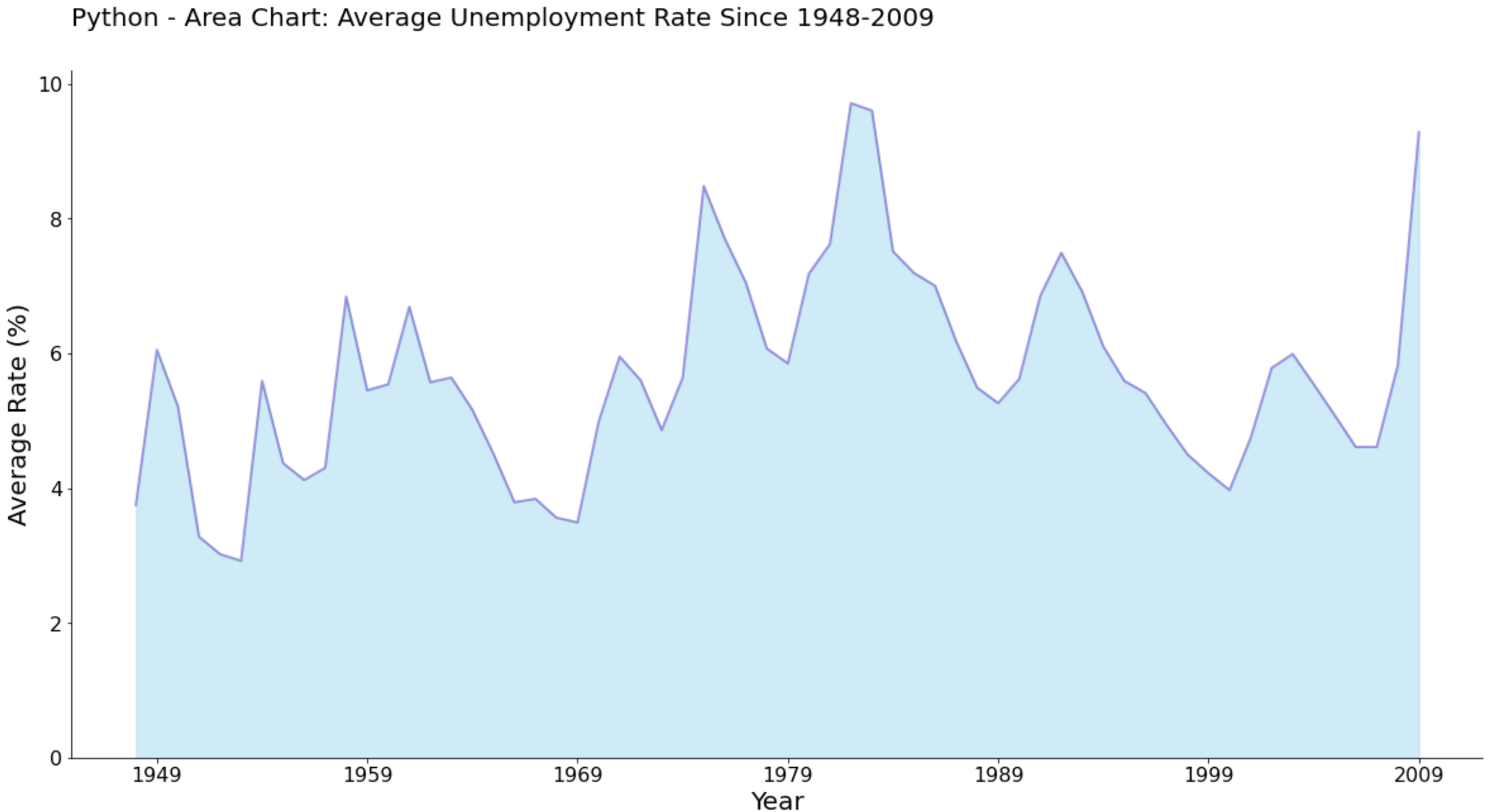
	Series id	Year	Period	Value
744	LNS14000000	2010	Jan	9.7
745	LNS14000000	2010	Feb	9.7

In [6]:
```
df.drop([744, 745], inplace=True)
```

In [7]:
```
avg = df.groupby('Year').mean()
avg.reset_index(level=0, inplace=True)
avg = avg.round({'Value': 2})
```

In [8]:
```
plt.rcParams['figure.figsize'] = [20,10]
fig, ax = plt.subplots()

plt.fill_between(avg['Year'], avg['Value'], color="skyblue", alpha=0.4)
plt.plot(avg['Year'], avg['Value'], color="Slateblue", alpha=0.6, linewidth=2)

plt.tick_params(labelsize=16)
plt.xticks([1949, 1959, 1969, 1979, 1989, 1999, 2009])
plt.suptitle("Python - Area Chart: Average Unemployment Rate Since 1948-2009",
```

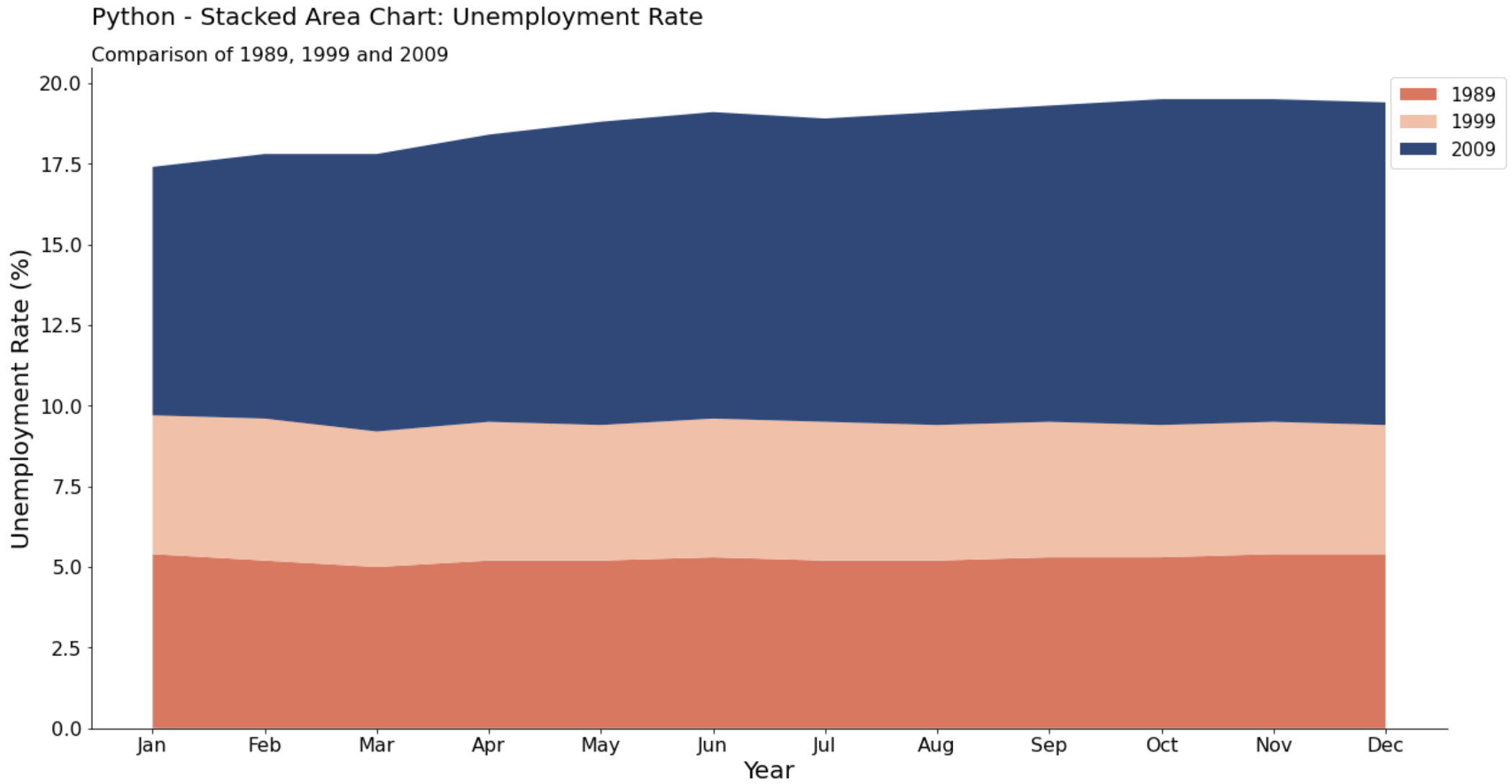
```
            size=20, x=.125, y=.95,horizontalalignment='left', verticalalignment='top')

plt.xlabel('Year', size=20)
plt.ylabel('Average Rate (%)', size=20)
plt.ylim(bottom=0)

right_side = ax.spines["right"]
right_side.set_visible(False)
top = ax.spines["top"]
top.set_visible(False)

plt.show()
```



In [9]:

```
y_1999 = df[df['Year']==1999]
y_2009 = df[df['Year']==2009]
y_1989 = df[df['Year']==1989]
```

In [25]:
```
plt.rcParams['figure.figsize'] = [20,10]
fig, ax = plt.subplots()

plt.stackplot(y_1989['Period'], y_1989['Value'],y_1999['Value'], y_2009['Value'], labels=['1989','1999','2009'],
              colors =['#D87860', '#f0c0a8', '#304878'])

plt.tick_params(labelsize=16)
plt.suptitle("Python - Stacked Area Chart: Unemployment Rate",
             size=20, x=.125, y=.95,horizontalalignment='left', verticalalignment='top')

plt.title("Comparison of 1989, 1999 and 2009", size=16, loc='left')
plt.xlabel('Year', size=20)
plt.ylabel('Unemployment Rate (%)', size=20)
plt.ylim(bottom=0)

right_side = ax.spines["right"]
right_side.set_visible(False)
top = ax.spines["top"]
top.set_visible(False)
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper right', borderpad=0.5, fontsize=15)

plt.show()
```



In [11]:

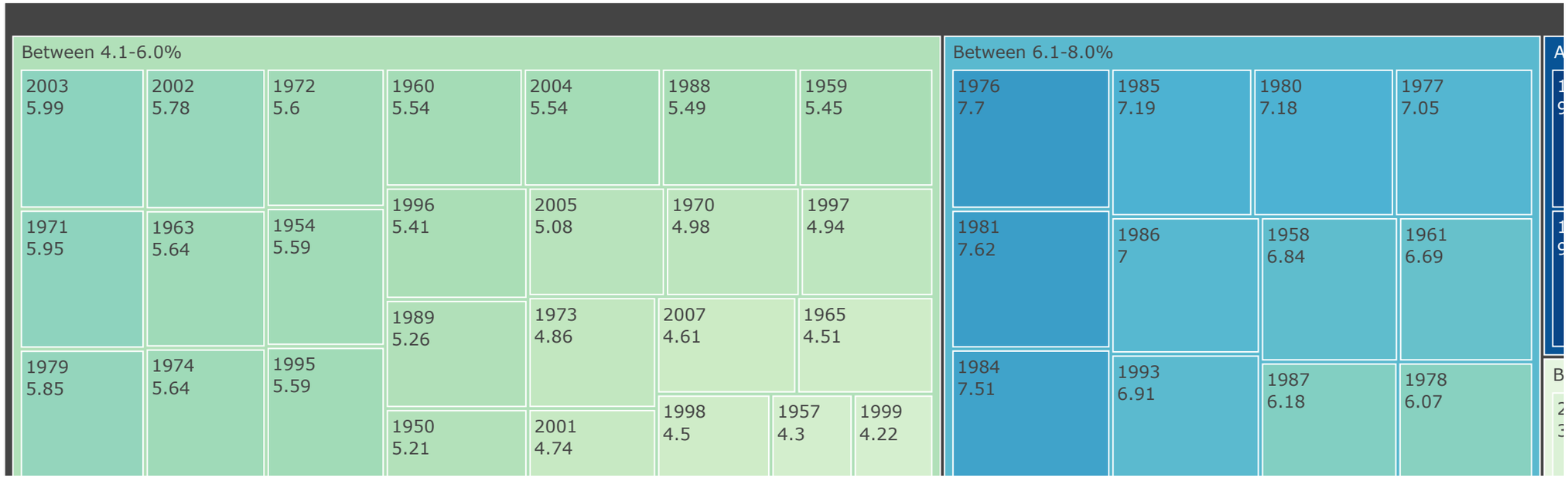
```
#binning method for confidence of fire.
bins = [2.5,4.0,6.0,8.0,11]
labels = ['Between 2.5-4.0%', 'Between 4.1-6.0%', 'Between 6.1-8.0%', 'Above 8.1%']
avg['Threshold'] = pd.cut(avg['Value'], bins=bins, labels=labels)
avg['Threshold'].fillna('Between 2.5-4.0%', inplace=True)
```

In [12]:
```
avg.rename(columns={"Value": "YearlyAverage"}, inplace=True)
```

In [13]:
```
fig = px.treemap(avg,
                 path=['Threshold', 'Year'],
                 values='YearlyAverage',
                 color='YearlyAverage',
                 color_continuous_scale='GnBu')

fig.update_layout(
    uniformtext=dict(minsize=10, mode='show'),
    margin = dict(t=50, l=25, r=25, b=25),
    title_text = 'Python - Tree Map: Average Unemployment Rate Distribution from 1948-2009',
    title_font_family="Arial",
    title_font_size = 22,
    title_font_color="black",
    title_x=0.03,
)
fig.data[0].textinfo = 'label+text+value'
fig.layout.hovermode = False
fig.show()
```

```
C:\Users\bibek\anaconda3\envs\my_env\lib\site-packages\plotly\express\_core.py:1637: FutureWarning: The frame.append method is deprecated and will be removed f
rom pandas in a future version. Use pandas.concat instead.
  df_all_trees = df_all_trees.append(df_tree, ignore_index=True)
C:\Users\bibek\anaconda3\envs\my_env\lib\site-packages\plotly\express\_core.py:1637: FutureWarning: The frame.append method is deprecated and will be removed f
rom pandas in a future version. Use pandas.concat instead.
  df_all_trees = df_all_trees.append(df_tree, ignore_index=True)
```

Python - Tree Map: Average Unemployment Rate Distribution from 1948-2009



R Script

Assignment 3.2: Tree Maps, Area Chart and Stacked Area Chart

DSC640

Taniya Adhikari

In [36]:
```
library(ggplot2)
library(readxl)
library(scales)
library(plyr)
library(dplyr)
```

In [3]:
```
df <- read.csv("unemployement-rate-1948-2010.csv")
head(df)
```

Series.id	Year	Period	Value
LNS14000000	1948	M01	3.4
LNS14000000	1948	M02	3.8
LNS14000000	1948	M03	4.0
LNS14000000	1948	M04	3.9
LNS14000000	1948	M05	3.5
LNS14000000	1948	M06	3.6

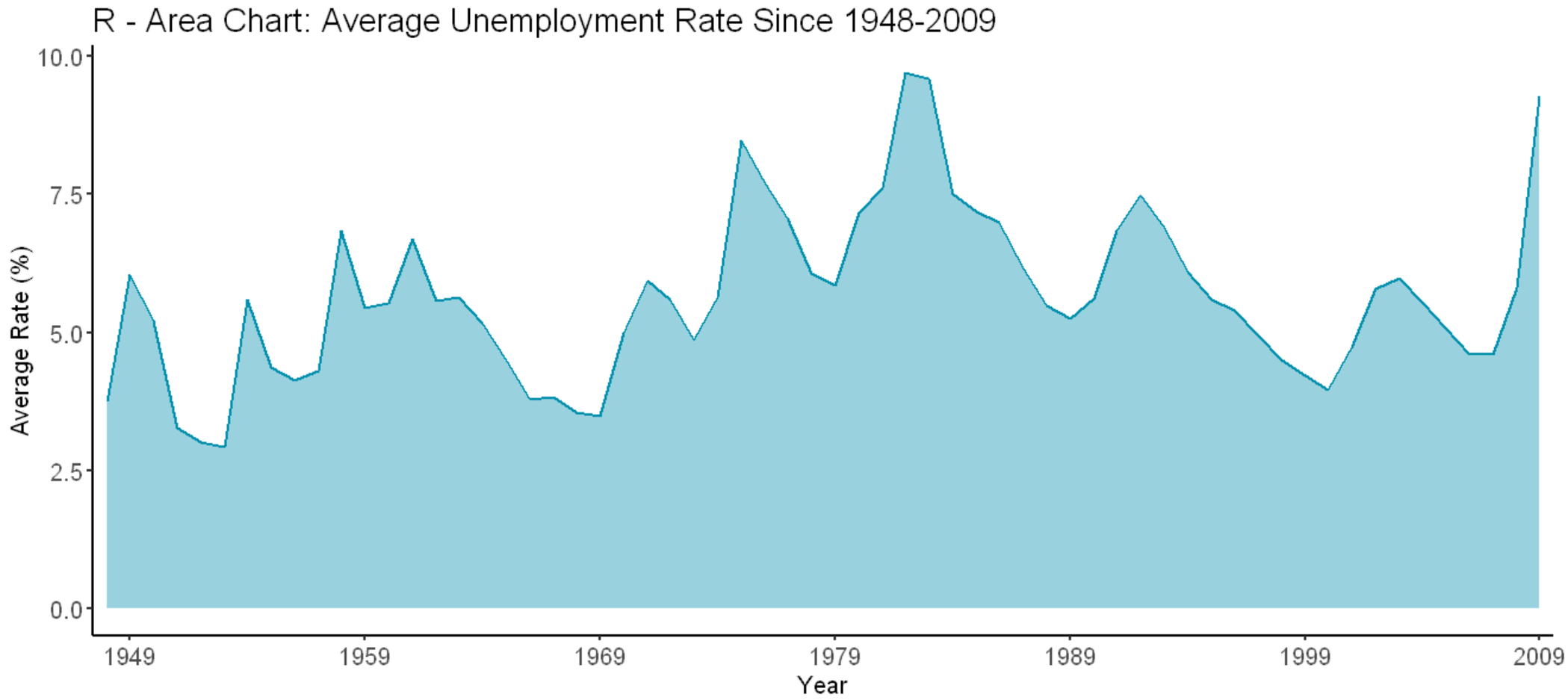
In [7]:
```
df<-df[!(df$Year==2010),]
```

In [17]:
```
avg = ddply(df, .(Year), summarize,  Average=mean(Value))
head(avg)
```

Year	Average
1948	3.750000
1949	6.050000
1950	5.208333
1951	3.283333
1952	3.025000
1953	2.925000

In [249…

```
options(repr.plot.width =11, repr.plot.height =5)
ggplot(avg, aes(x=Year, y=Average)) +
  geom_area( fill='#008FAD', alpha=0.4) +
  geom_line(size = .6, color='#008FAD') +
  theme_classic() +
  theme(text = element_text(family="sans",size =12, color="black"), element_line(size = .6),
        plot.title = element_text(size = 16), axis.text.x = element_text(size=12),
        axis.text.y = element_text(size=12)) +
  expand_limits(y = c(0, NA)) +
  scale_x_discrete(name ="Year", limits=c(1949,1959,1969, 1979,1989, 1999, 2009)) +
  ylab("Average Rate (%)") +
  ggtitle("R - Area Chart: Average Unemployment Rate Since 1948-2009")
```



In [195…

```
mygroups<-df[(df$Year==1989 |df$Year==1999 | df$Year==2009),]
mygroups$Period <- as.character(mygroups$Period)
```

In [196…

```
mygroups[mygroups == "M01"] <- 1
mygroups[mygroups == "M02"] <- 2
mygroups[mygroups == "M03"] <- 3
mygroups[mygroups == "M04"] <- 4
mygroups[mygroups == "M05"] <- 5
mygroups[mygroups == "M06"] <- 6
mygroups[mygroups == "M07"] <- 7
```

```
mygroups[mygroups == "M08"] <- 8
mygroups[mygroups == "M09"] <- 9
mygroups[mygroups == "M10"] <- 10
mygroups[mygroups == "M11"] <- 11
mygroups[mygroups == "M12"] <- 12
```

In [197…

```
mygroups$Period = as.factor(mygroups$Period)
```

In [198…

```
mygroups$Year <- as.factor(mygroups$Year)
```

In [199…

```
head(mygroups)
```

	Series.id	Year	Period	Value
493	LNS14000000	1989	1	5.4
494	LNS14000000	1989	2	5.2
495	LNS14000000	1989	3	5.0
496	LNS14000000	1989	4	5.2
497	LNS14000000	1989	5	5.2
498	LNS14000000	1989	6	5.3

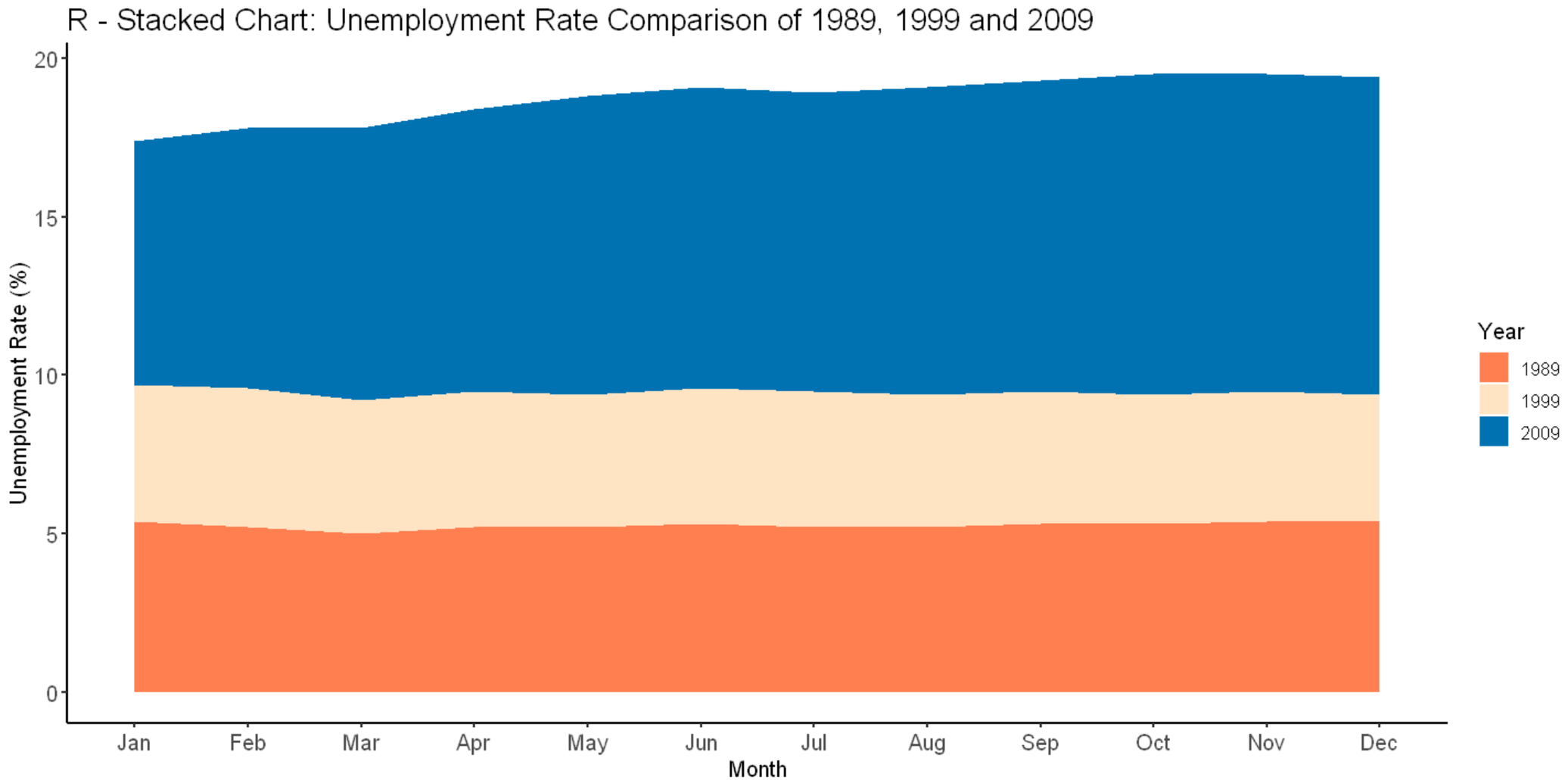
In [200…

```
# Give a specific order:
mygroups$Year <- factor(mygroups$Year, levels=c("2009", "1999", "1989"))
```

In [248…

```
options(repr.plot.width =12, repr.plot.height =6)
level_order <- c('1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12')
ggplot(mygroups, aes(x=factor(Period,level=level_order), y=Value)) +
  geom_area(aes(group=Year, fill =Year)) +
  scale_fill_manual(values = c("#0072B2","bisque","coral")) +
  guides(fill = guide_legend(reverse=TRUE)) +
  theme_classic() +
  theme(text = element_text(family="sans",size =12, color="black"), element_line(size = .6),
        plot.title = element_text(size = 16), axis.text.x = element_text(size=12),
        axis.text.y = element_text(size=12)) +
  expand_limits(y = c(0, NA)) +
```

```
scale_x_discrete(labels=c("1" = "Jan", "2" = "Feb", "3" = "Mar", "4" ="Apr", "5" ="May", "6" ="Jun",
                          "7" = "Jul", "8" = "Aug", "9" = "Sep", "10" = "Oct", "11" = "Nov", "12" = "Dec")) +
ylab("Unemployment Rate (%)") + xlab("Month") +
ggtitle("R - Stacked Chart: Unemployment Rate Comparison of 1989, 1999 and 2009")
```



In [212…

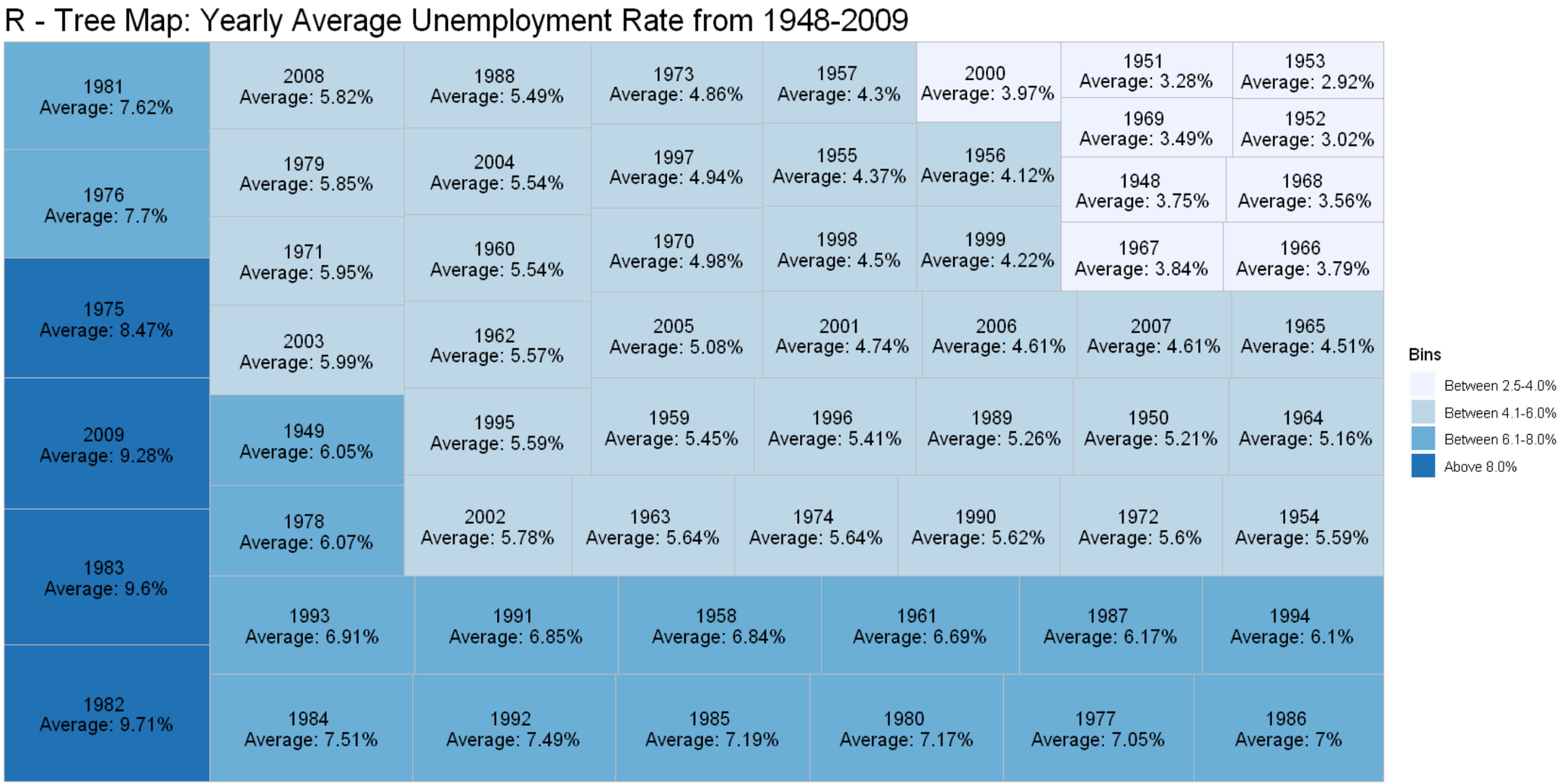
```
avg$Bins <-""
```

In [218…
```
avg[avg$Average < 4.0, "Bins"] <- "Between 2.5-4.0%"
avg[avg$Average > 4.0 & avg$Average < 6.0, "Bins"] <- "Between 4.1-6.0%"
avg[avg$Average > 6.0 & avg$Average < 8.0, "Bins"] <- "Between 6.1-8.0%"
avg[avg$Average > 8.0, "Bins"] <- "Above 8.0%"
```

In [232…
```
# Give a specific order:
avg$Bins <- factor(avg$Bins, levels=c("Between 2.5-4.0%", "Between 4.1-6.0%", "Between 6.1-8.0%","Above 8.0%"))
```

In [229…
```
library(treemapify)
```

In [254…
```
options(repr.plot.width =14, repr.plot.height =7)
ggplot(avg, aes(area =Average, fill =Bins,
                label =paste0(Year, " ", "\n", "Average: ",round(Average, digits = 2), "%"), subgroup=Year)) +
  geom_treemap() +
  theme(plot.title = element_text(size = 20))+
  geom_treemap_text(colour = "black",
                    place = "centre",
                    size = 12) +
  scale_fill_brewer(palette = "Blues") +
  ggtitle("R - Tree Map: Yearly Average Unemployment Rate from 1948-2009")
```


R - Tree Map: Yearly Average Unemployment Rate from 1948-2009
1981 Average: 7.62%
2008 Average: 5.82%
1988 Average: 5.49%
1973 Average: 4.86%
1957 Average: 4.3%
2000 Average: 3.97%
1951 Average: 3.28%
1953 Average: 2.92%
1969 Average: 3.49%
1952 Average: 3.02%
1976 Average: 7.7%
1979 Average: 5.85%
2004 Average: 5.54%
1997 Average: 4.94%
1955 Average: 4.37%
1956 Average: 4.12%
1948 Average: 3.75%
1968 Average: 3.56%
1971 Average: 5.95%
1960 Average: 5.54%
1970 Average: 4.98%
1998 Average: 4.5%
1999 Average: 4.22%
1967 Average: 3.84%
1966 Average: 3.79%
1975 Average: 8.47%
2003 Average: 5.99%
1962 Average: 5.57%
2005 Average: 5.08%
2001 Average: 4.74%
2006 Average: 4.61%
2007 Average: 4.61%
1965 Average: 4.51%
2009 Average: 9.28%
1949 Average: 6.05%
1995 Average: 5.59%
1959 Average: 5.45%
1996 Average: 5.41%
1989 Average: 5.26%
1950 Average: 5.21%
1964 Average: 5.16%
1983 Average: 9.6%
1978 Average: 6.07%
2002 Average: 5.78%
1963 Average: 5.64%
1974 Average: 5.64%
1990 Average: 5.62%
1972 Average: 5.6%
1954 Average: 5.59%
1993 Average: 6.91%
1991 Average: 6.85%
1958 Average: 6.84%
1961 Average: 6.69%
1987 Average: 6.17%
1994 Average: 6.1%
1982 Average: 9.71%
1984 Average: 7.51%
1992 Average: 7.49%
1985 Average: 7.19%
1980 Average: 7.17%
1977 Average: 7.05%
1986 Average: 7%
Bins
Between 2.5-4.0%
Between 4.1-6.0%
Between 6.1-8.0%
Above 8.0%

In [ ]:

In [ ]:

Assignment3 - Power BI Desktop

Search

Sign in

File Home Insert Modeling View Help

Paste Cut Copy Format painter — Clipboard

Get data Excel workbook Power BI datasets SQL Server Enter data Dataverse Recent sources — Data

Transform data Refresh — Queries

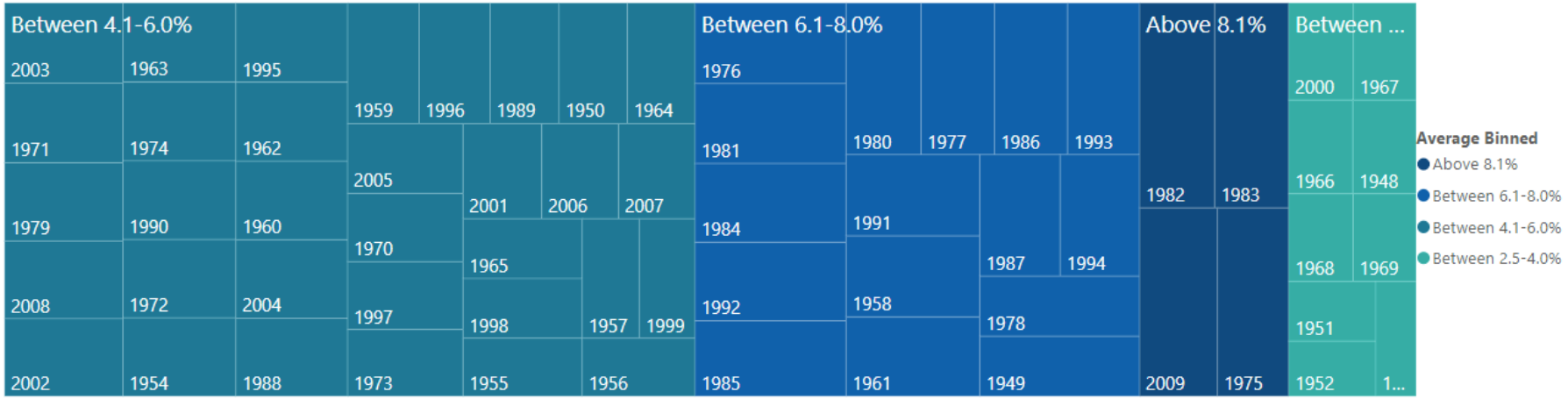
New visual Text box More visuals — Insert

New measure Quick measure — Calculations

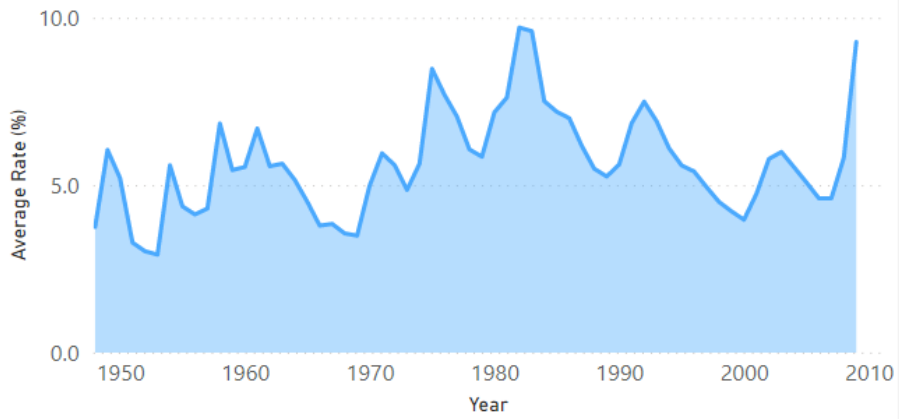
Sensitivity — Sensitivity

Publish — Share

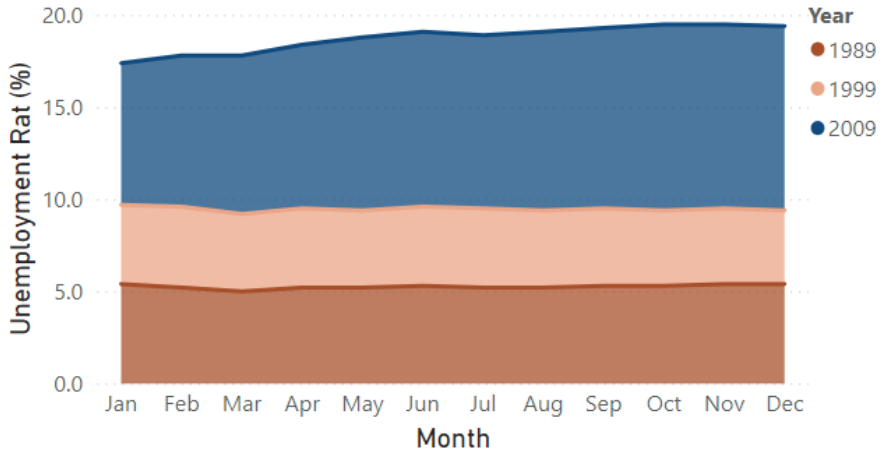
PowerBI - Tree Map: Average Unemployment Rate Distribution from 1948-2009



PowerBI - Area Chart: Average Employment Rate since 1948-2009



PowerBi - Stacked Area Chart: Unemployment Rate Compari...



Filters

Page 1 Page 2 Page 3 Page 4 +