

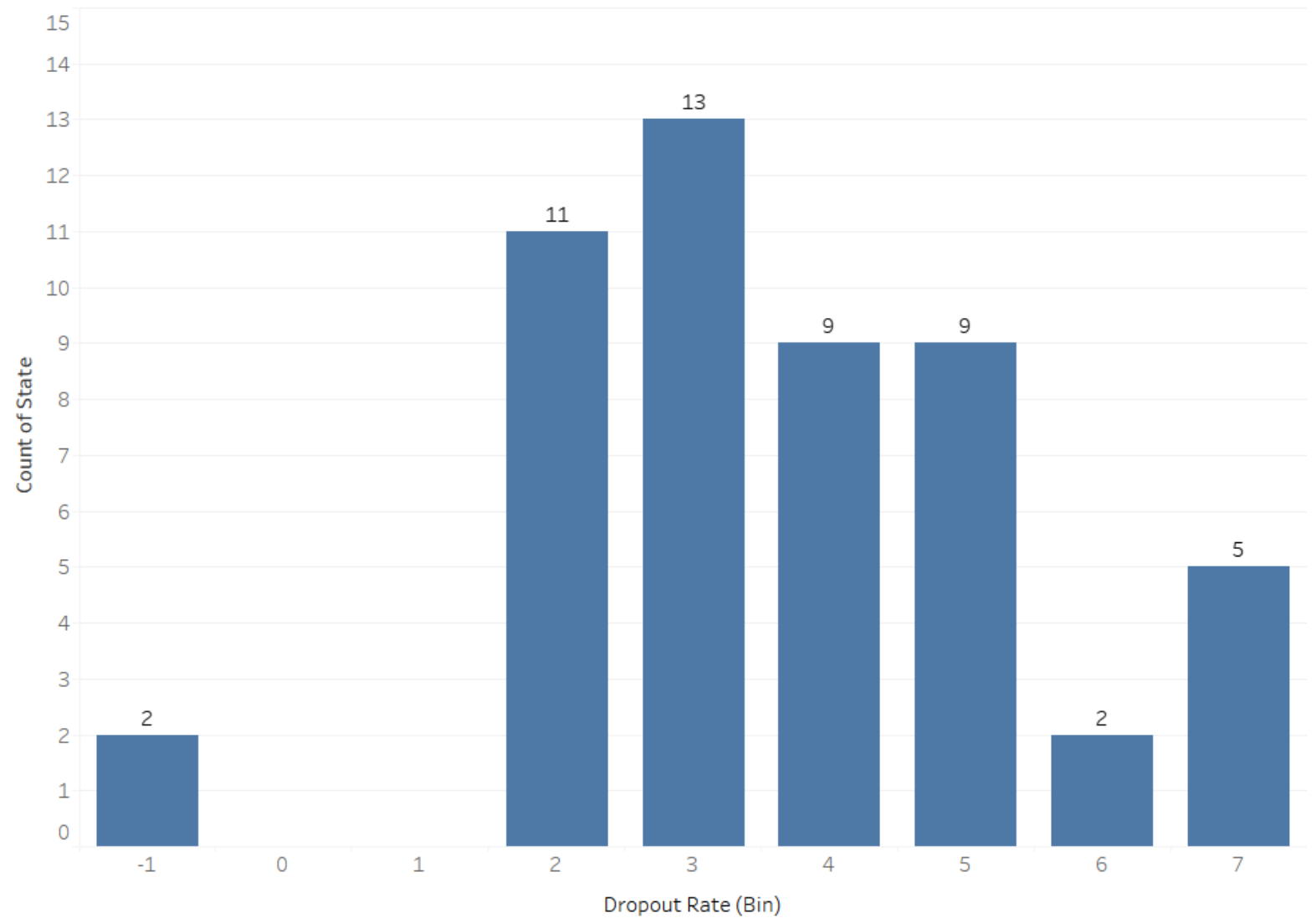
# Contents

Visualization	Page
Tableau Histogram	2
Tableau Box Plot	3
Tableau Bullet Chart	4
Tableau Above/Below Bar Chart	5
Python Histogram	6
Python Box Plot	7
Python Bullet Chart	8
Python Above/Below Chart	9
R Histogram	10
R Box Plot	11
R Bullet Chart	12
R Above/Below Chart	13

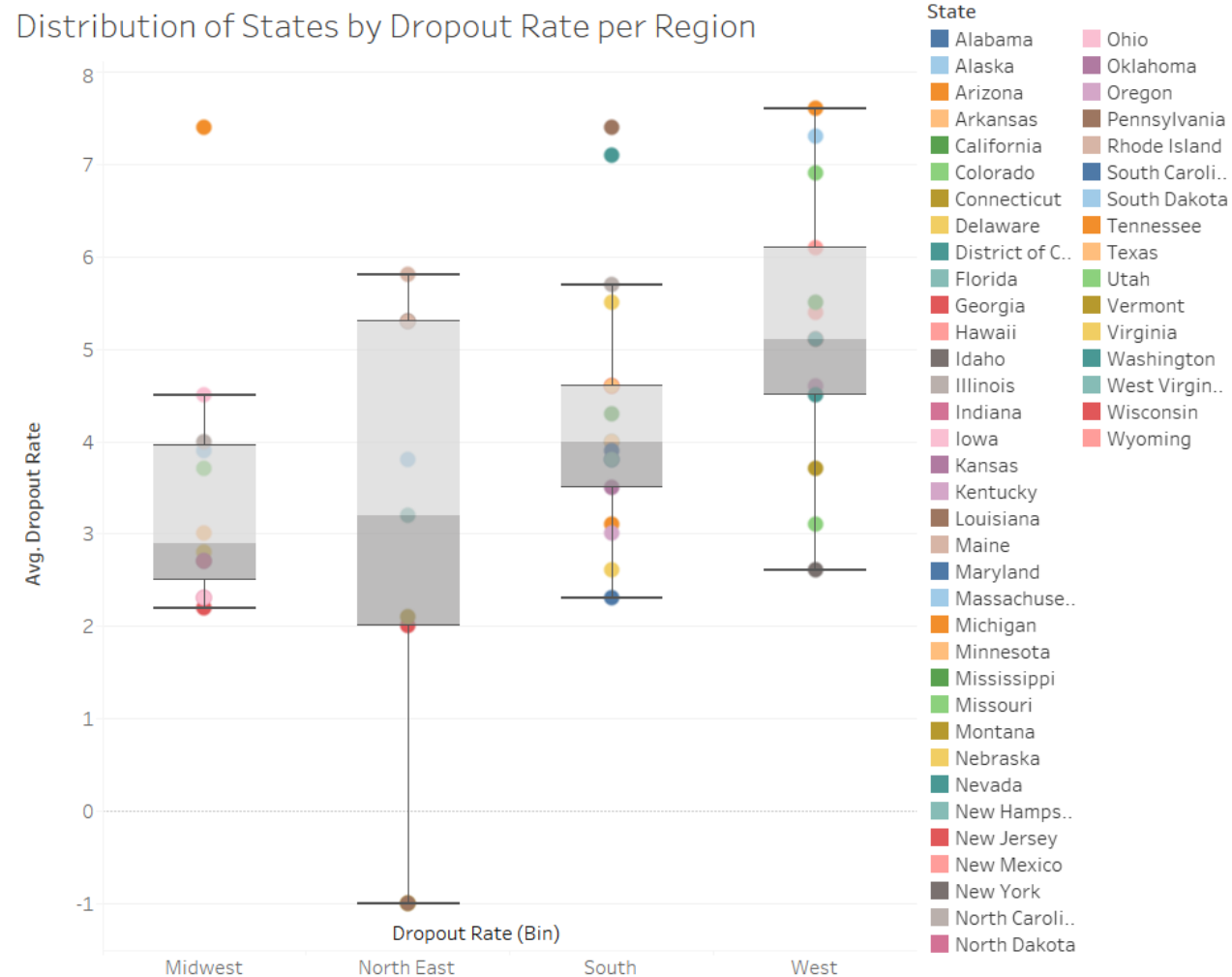


# Tableau – Histogram

Distribution of States by Dropout Rate

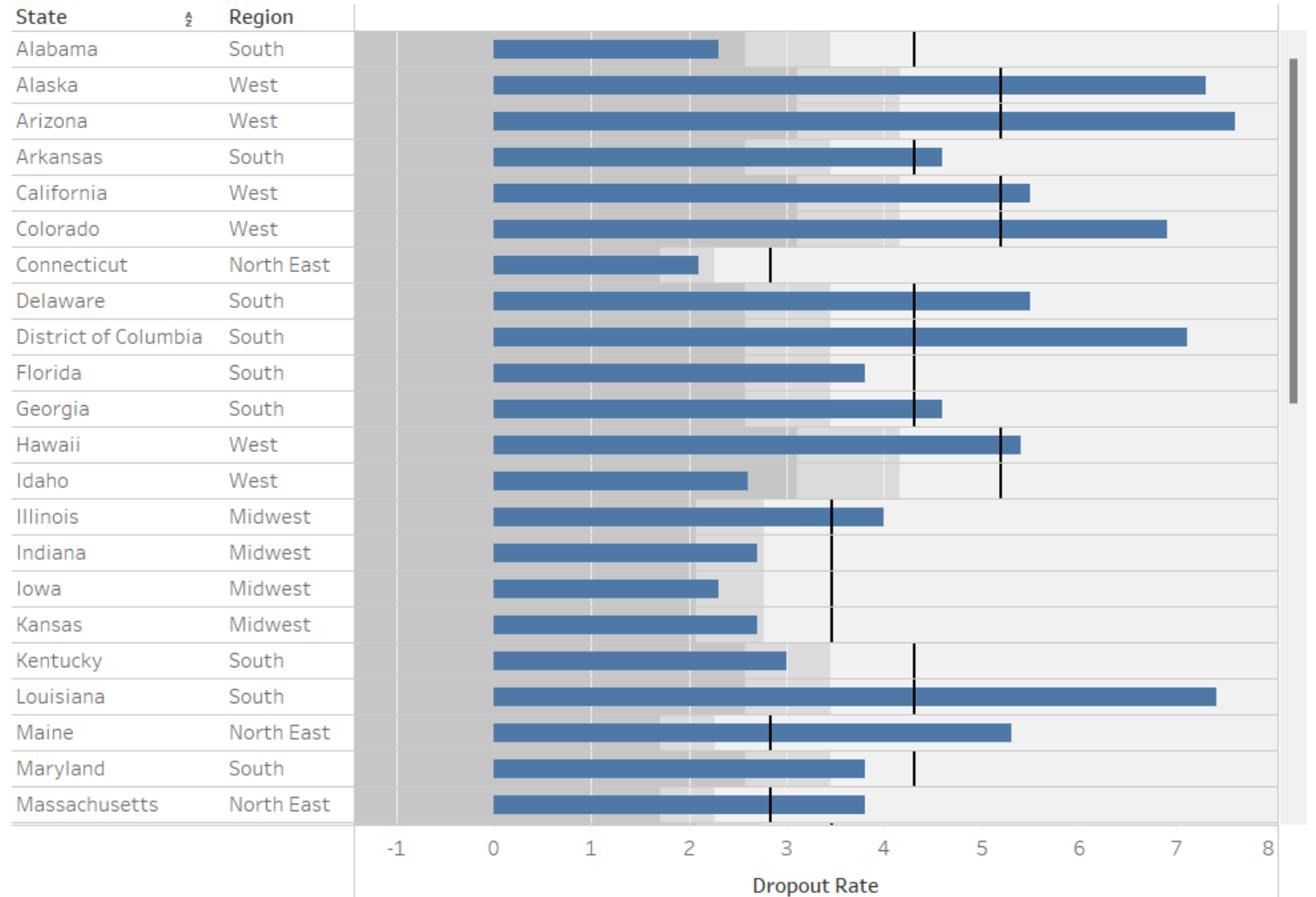


# Tableau – Box Plot



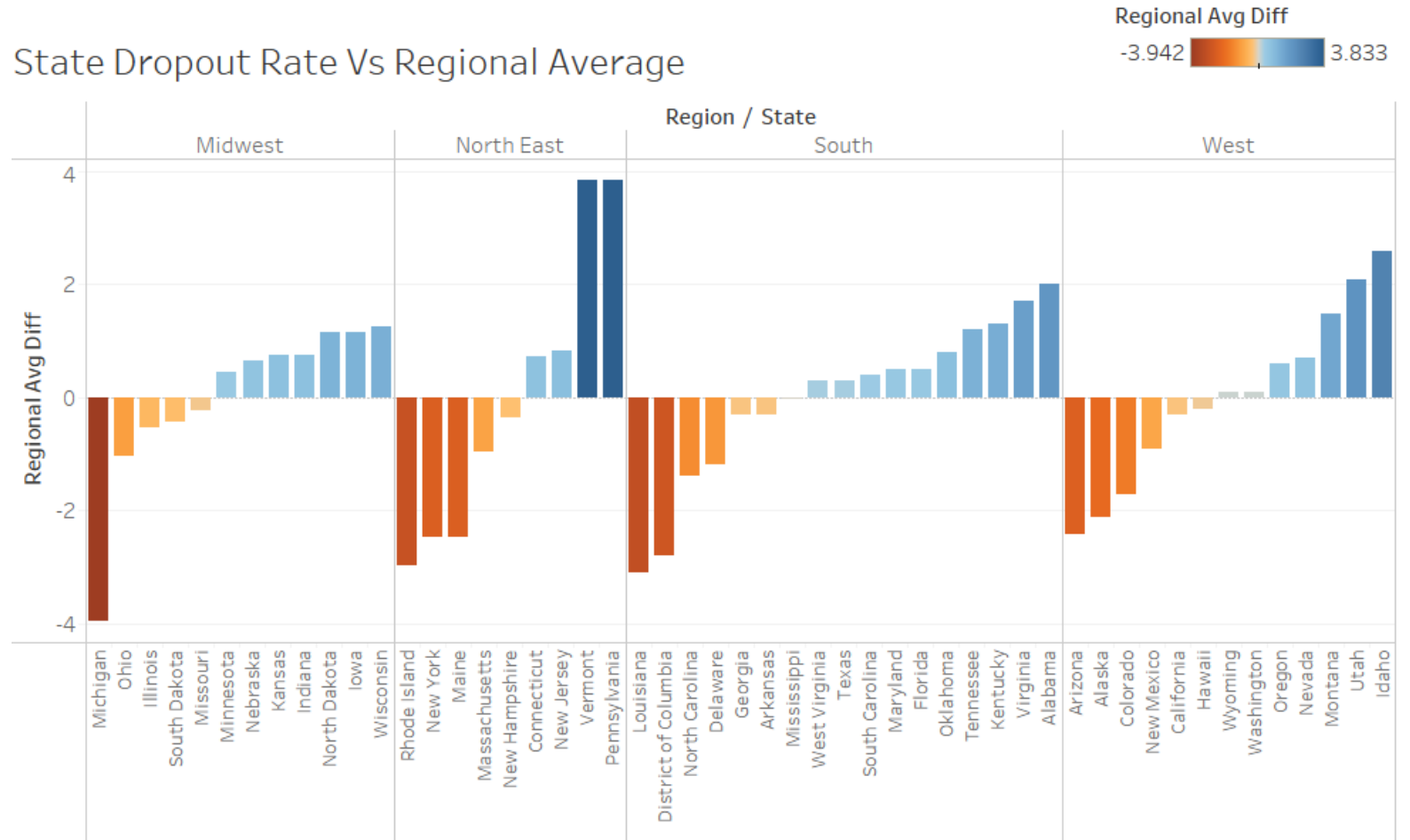
# Tableau – Bullet Chart

Dropout Rate by State  
Compared to Average Regional Dropout Rate



# Tableau – Above/Below Bar Chart

State Dropout Rate Vs Regional Average



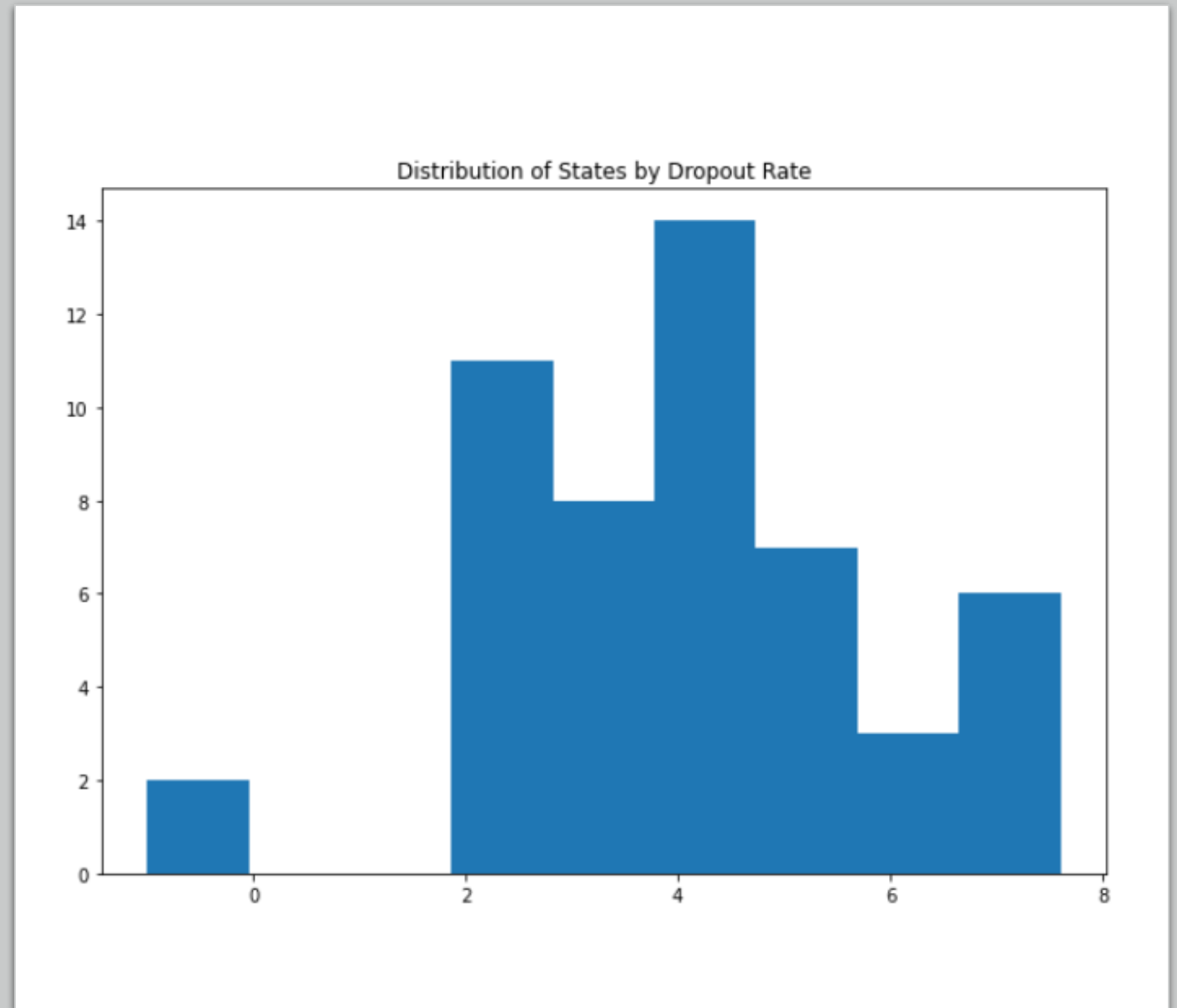
# Python – Histogram

## Code

```
import pandas as pd
from matplotlib.pyplot import figure

# Import data to dataframe
df =
pd.read_csv(r'C:\Users\jeric\OneDrive/Documents/classFiles/DSC640/Exercises/Week1
0_11/education.csv', index_col=False)

# Creating histogram
fig, ax = plt.subplots(figsize =(10, 7))
ax.hist(df['dropout_rate'], 9)
ax.set_title("Distribution of States by Dropout Rate")
```



# Python – Box Plot

## Code

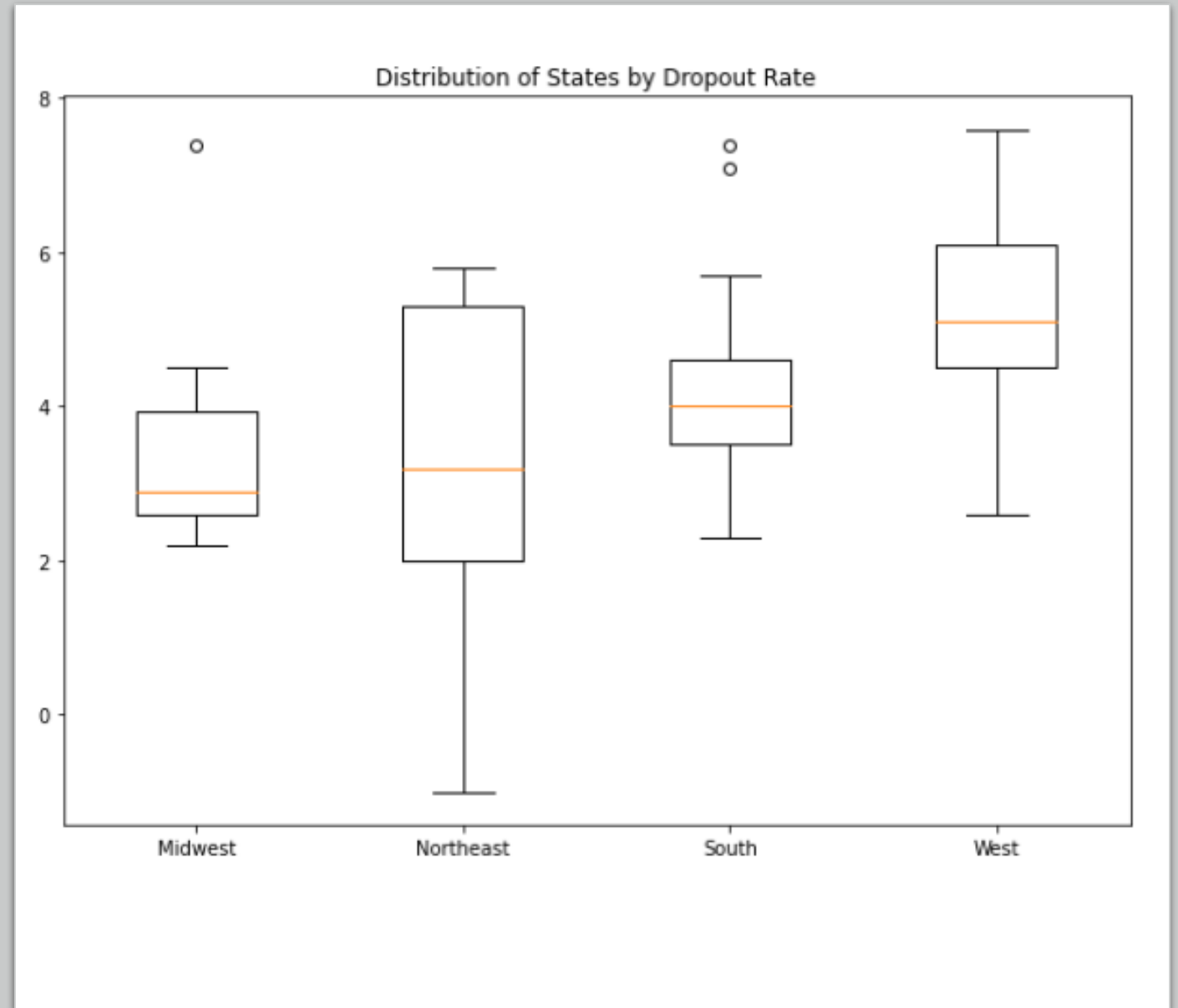
```
import pandas as pd
import matplotlib.pyplot as plt

# Import data to dataframe
df =
pd.read_csv(r'C:\Users\jeric\OneDrive\Documents\classFiles\DSC640\Exercises\Week10_11\education.csv',
index_col=False)

# Setting Groups and splitting dataframe
grouped = df.groupby(df.Region)
df_MW = grouped.get_group("Midwest")
df_NE = grouped.get_group("North East")
df_S = grouped.get_group("South")
df_W = grouped.get_group("West")

# Recombining groups
data = [df_MW['dropout_rate'], df_NE['dropout_rate'], df_S['dropout_rate'], df_W['dropout_rate']]

# Creating histogram
fig, ax = plt.subplots(figsize =(10, 7))
plt.boxplot(data, labels=['Midwest', 'Northeast', 'South', 'West'])
ax.set_title("Distribution of States by Dropout Rate")
```

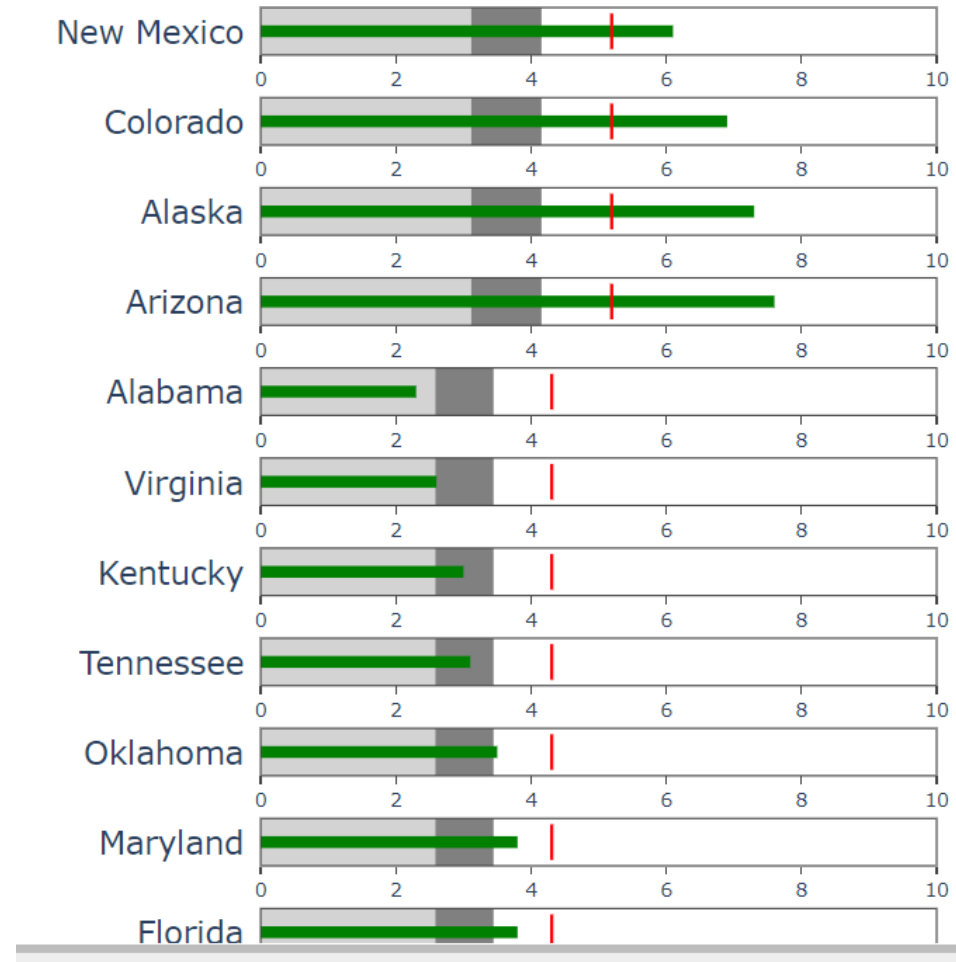


# Python – Bullet Chart

## Code

Code was too long to add here

Dropout Rate by State  
Compared to Average Regional Dropout Rate





# Python – Above/Below Chart

Code

```
import pandas as pd
import matplotlib.pyplot as plt
```

```
# Import data to dataframe
```

```
df =
pd.read_csv(r'C:\Users\jeric\OneDrive/Documents/classFiles/DSC640/Exercise
s/Week10_11/education.csv', index_col=False)
```

```
# Adding Regional Average
```

```
df['newname'] = df["Region"]+"-"+df["state"]
df['Region_avg'] = df.groupby(["Region"])[ 'dropout_rate'].transform('mean')
df['Region_avg_diff'] = df['Region_avg']-df['dropout_rate']
df = df.sort_values(by=["Region",'Region_avg_diff'], ascending=True)
```

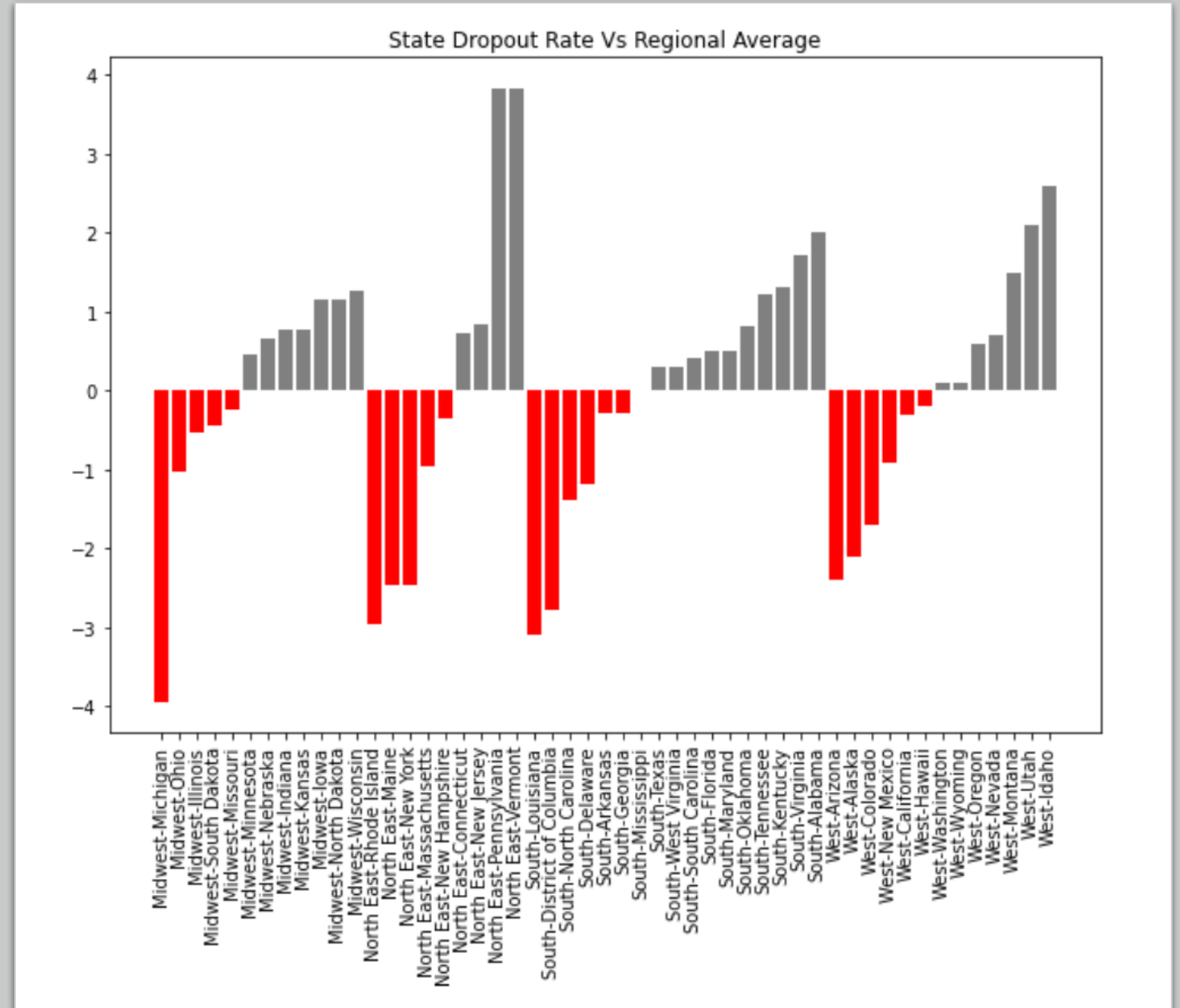
```
colors = ['r' if (bar < 0) else 'grey' for bar in df['Region_avg_diff']]
```

```
fig, ax = plt.subplots(figsize =(10, 7))
```

```
plt.xticks(rotation = 90)
```

```
ax.bar(df['newname'],df['Region_avg_diff'], color=colors)
```

```
ax.set_title("State Dropout Rate Vs Regional Average")
```



# R – Histogram

Code

```
# Histogram
```

```
library(ggplot2)
```

```
## Set the working directory
```

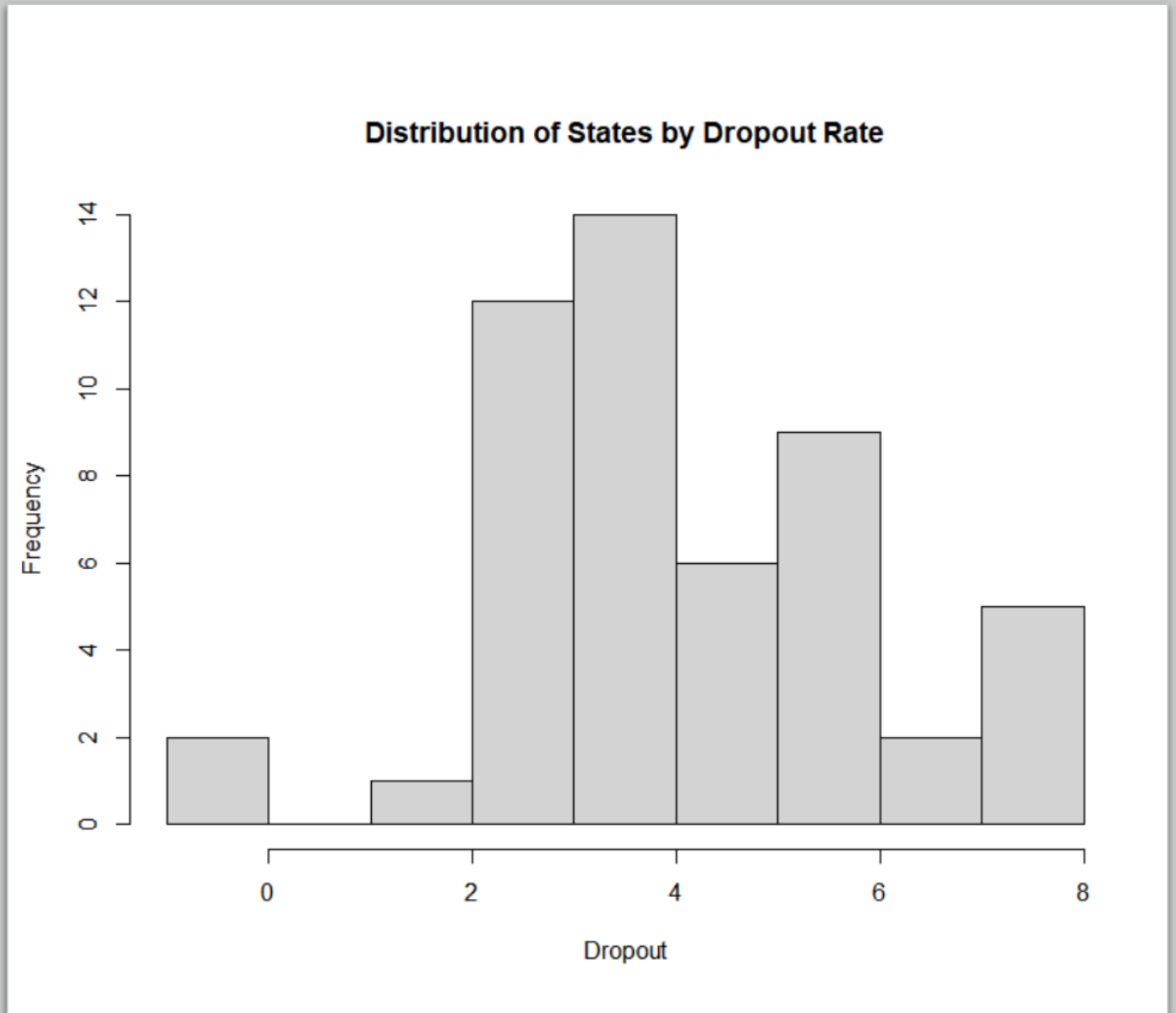
```
setwd('C:/Users/jeric/OneDrive/Documents/classFiles/DSC640/Exercises/Week10_11')
```

```
## Load the data to a dataframe
```

```
df <- read.csv("education.csv")
```

```
Dropout <- df$dropout_rate
```

```
hist(Dropout, main="Distribution of States by Dropout Rate")
```



# R – Box Plot

Code

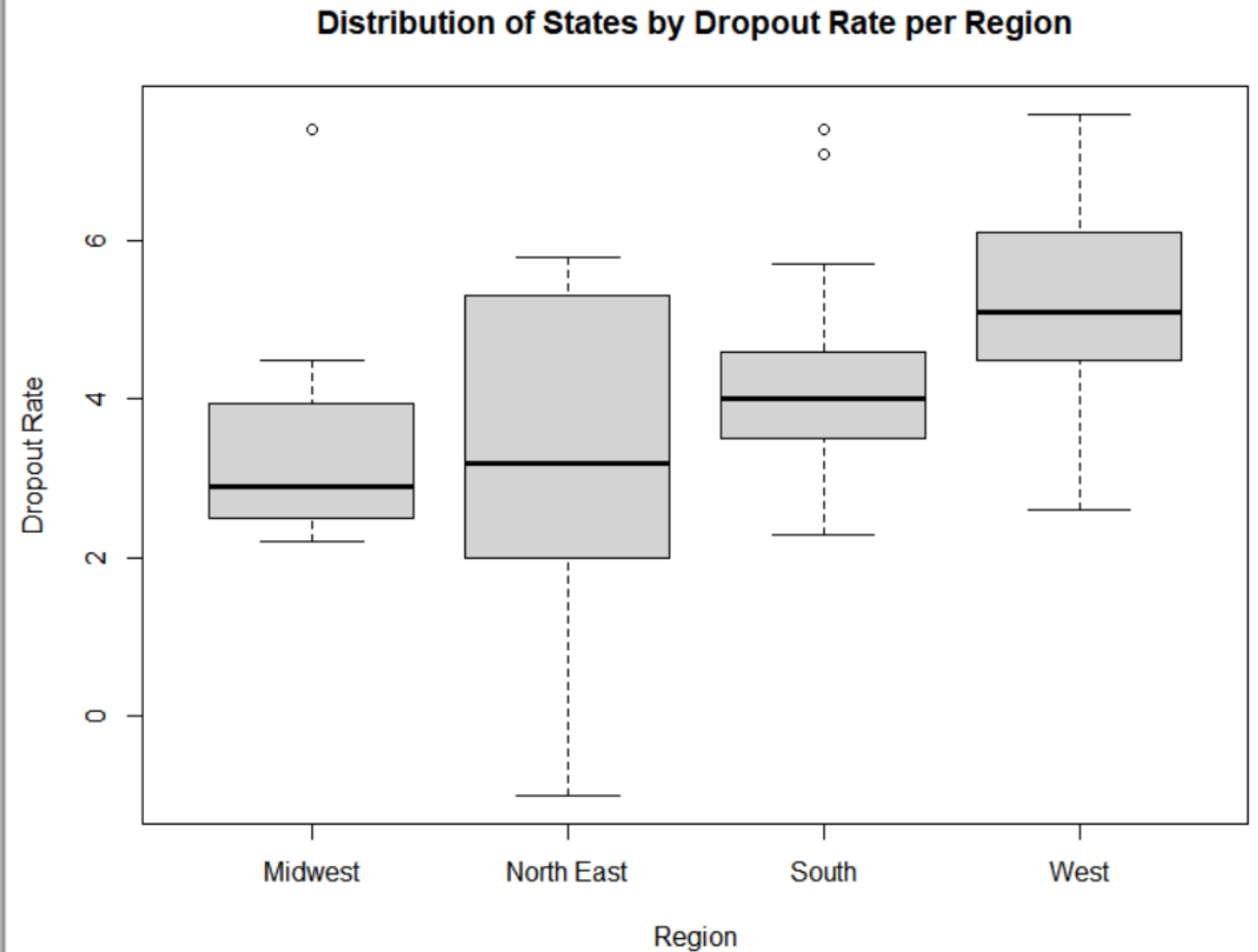
```
library(ggplot2)

## Set the working directory
setwd('C:/Users/jeric/OneDrive/Documents/classFiles/DSC640/Exercises/Week10_11')

## Load the data to a dataframe
df <- read.csv("education.csv")

Dropout <- df$dropout_rate

boxplot(dropout_rate~Region,data=df, main="Distribution of States by Dropout Rate per Region",
        xlab="Region", ylab="Dropout Rate")
```



# R – Bullet Chart

Code

# R – Above/Below Chart

Code

```
library(ggplot2)
```

```
library(dplyr)
```

```
## Set the working directory
```

```
setwd('C:/Users/jeric/OneDrive/Documents/classFiles/DSC640/Exercises/Week10_11')
```

```
## Load the data to a dataframe
```

```
df <- read.csv("education.csv")
```

```
# Add Regional Average Field
```

```
df <- df %>%
```

```
  group_by(Region) %>%
```

```
  mutate(Regional_avg = mean(dropout_rate))
```

```
# Add Regional Avg Diff Field
```

```
df$Regional_avg_diff <- df$Regional_avg - df$dropout_rate
```

```
barplot(df$Regional_avg_diff, names.arg=c(df$state), las=2, main="State  
Dropout Rate Vs Regional Average" )
```

