Assignment 6.2 Charts

DSC640

Taniya Adhikari

Tableau – Histogram

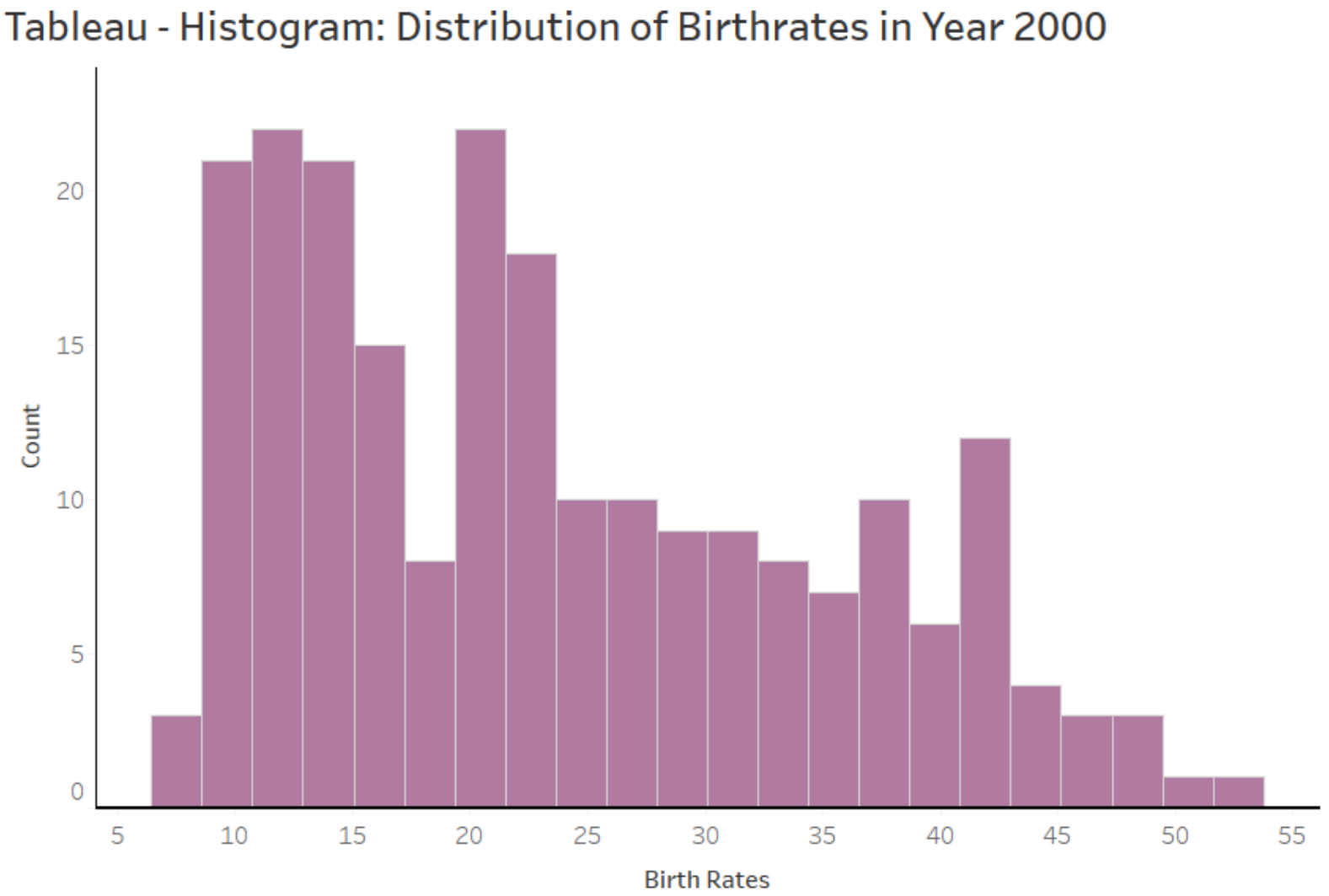


Tableau – Boxplot

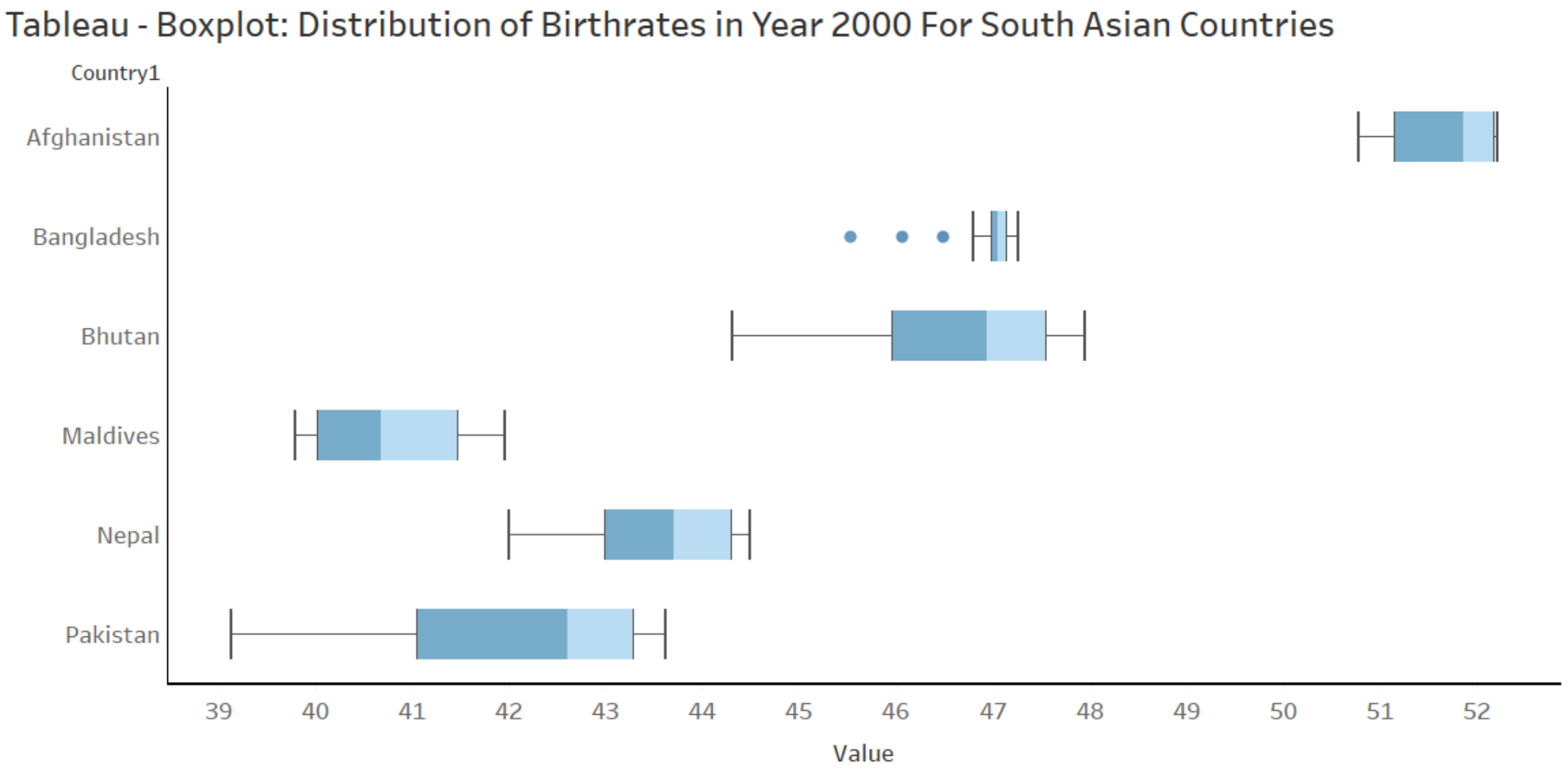


Tableau – Bullet Chart

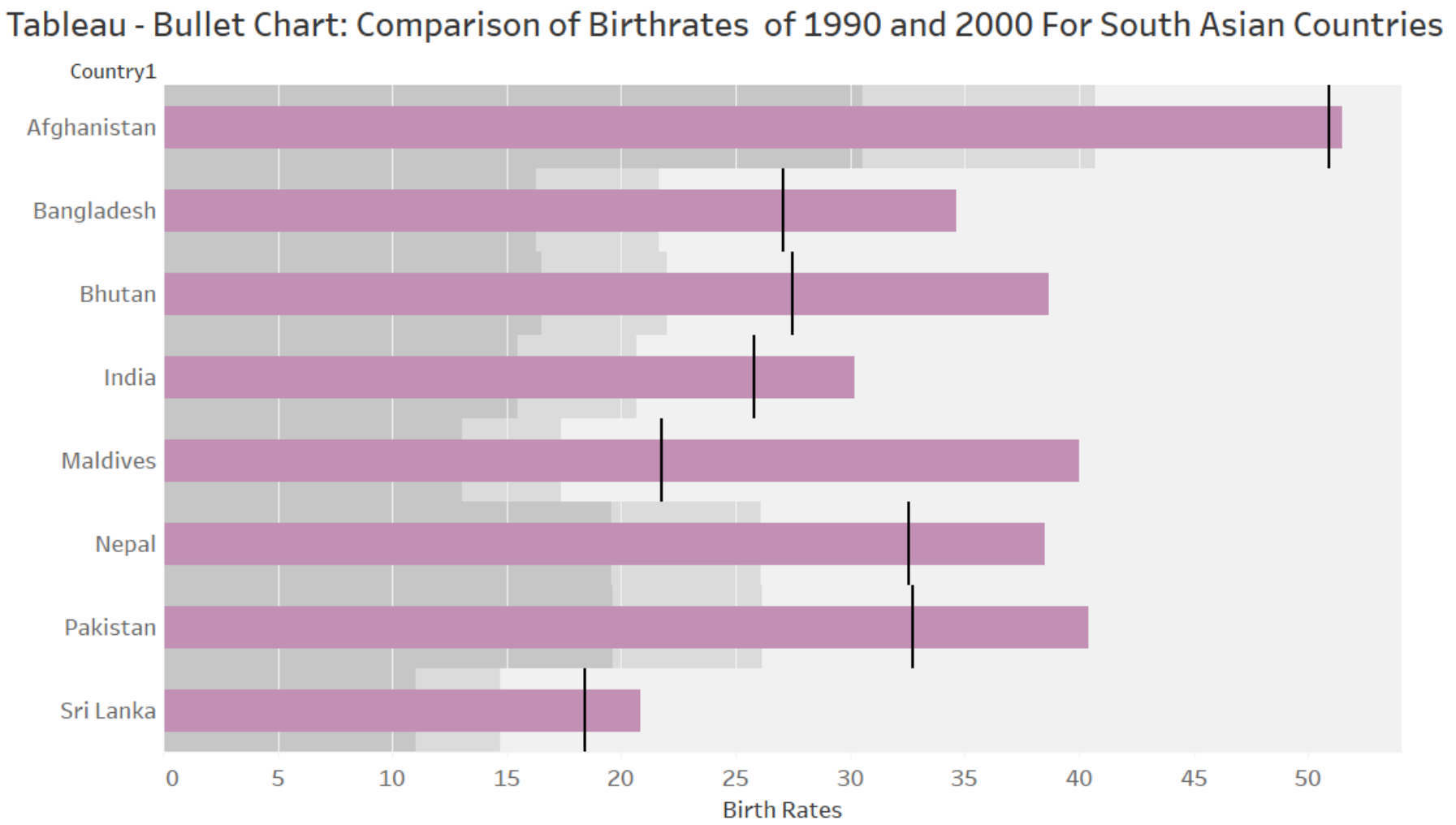
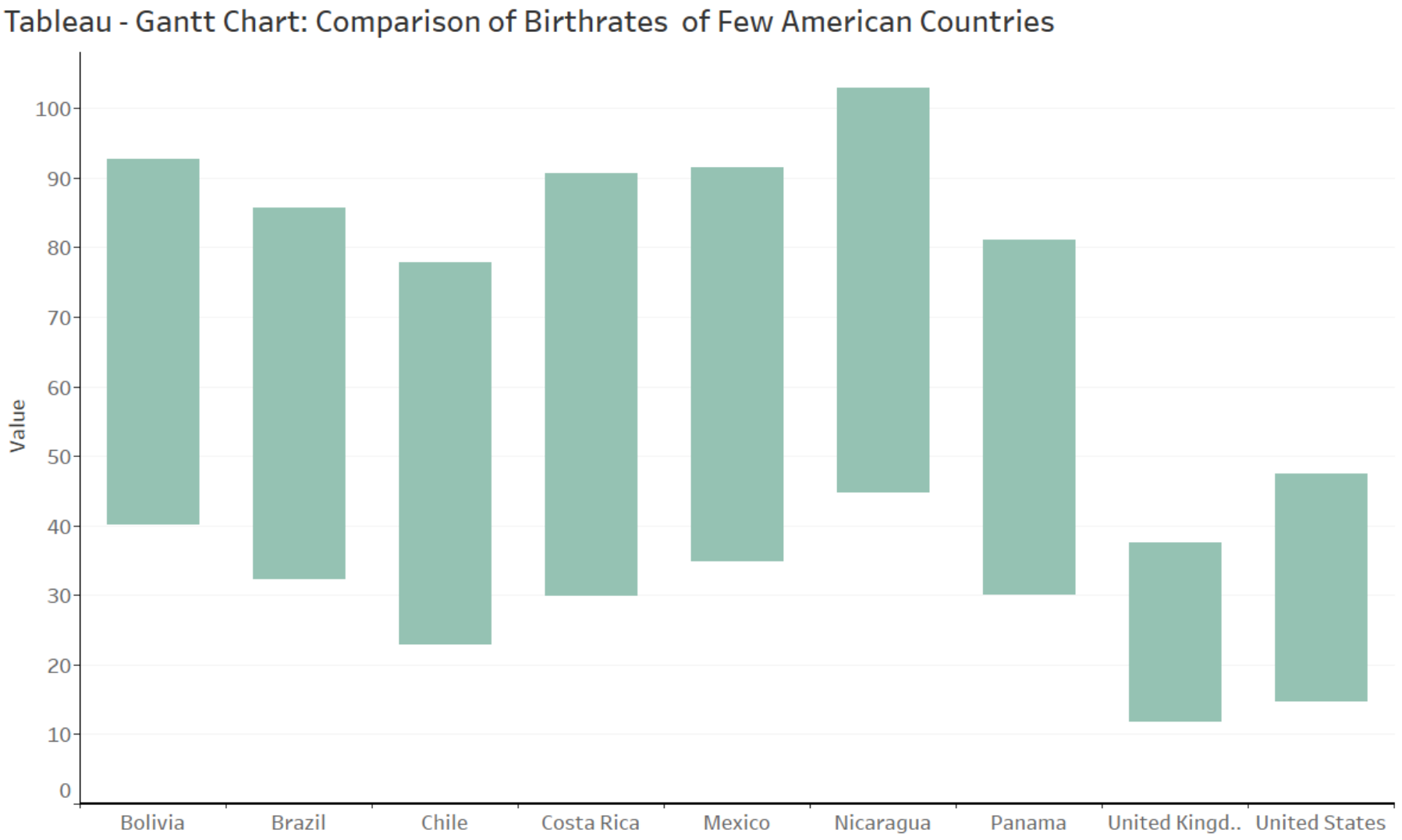
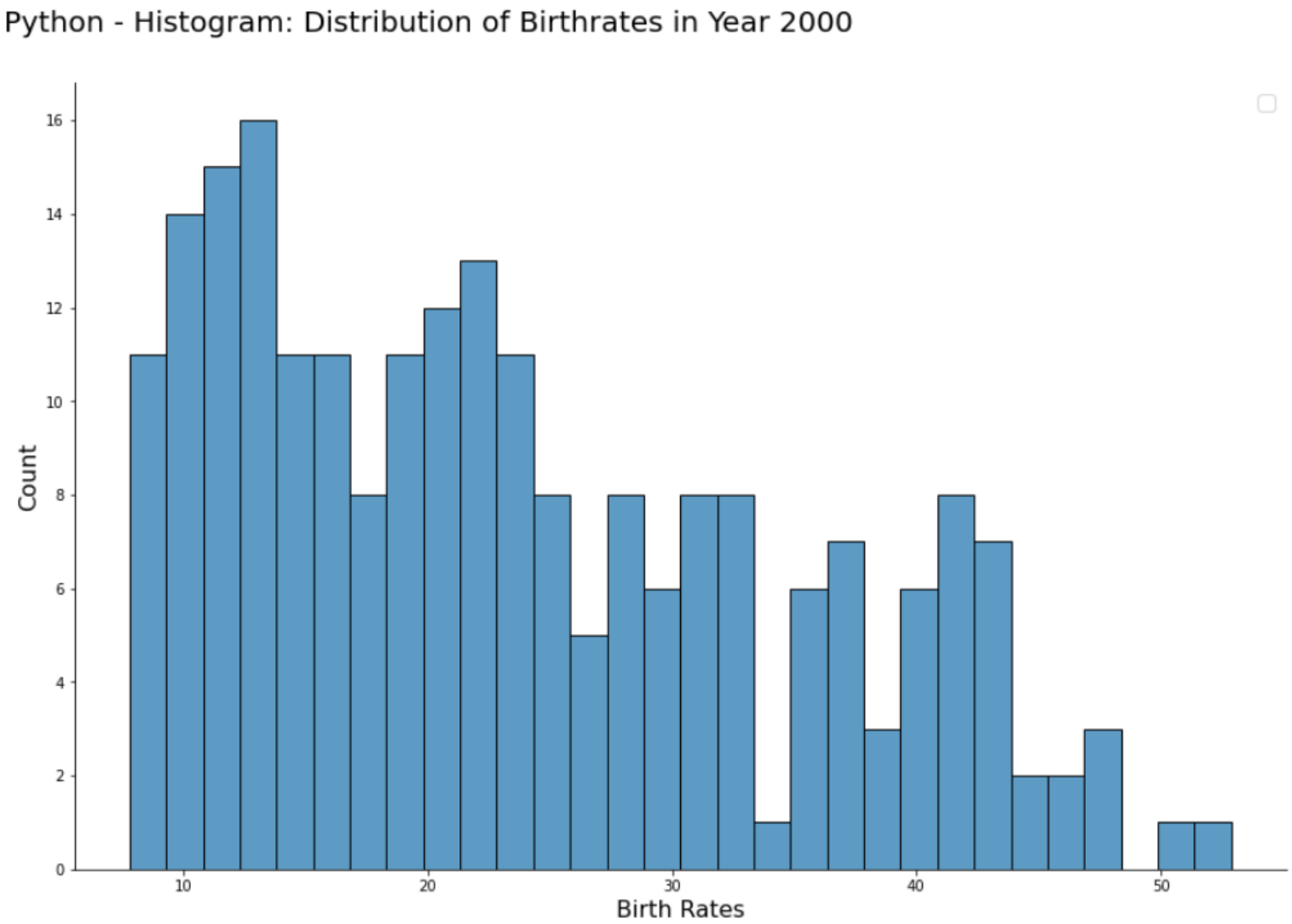


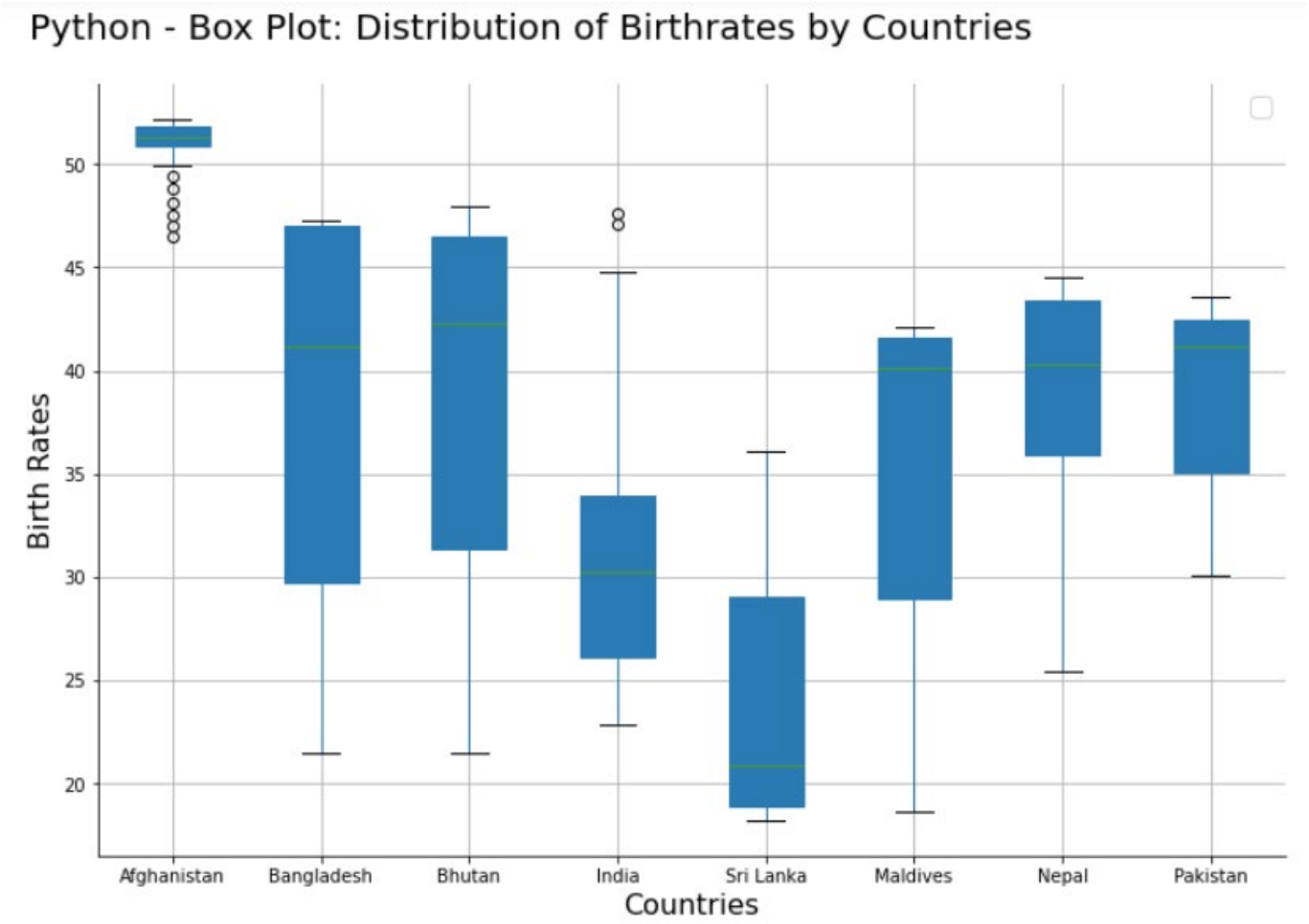
Tableau – Gantt Chart



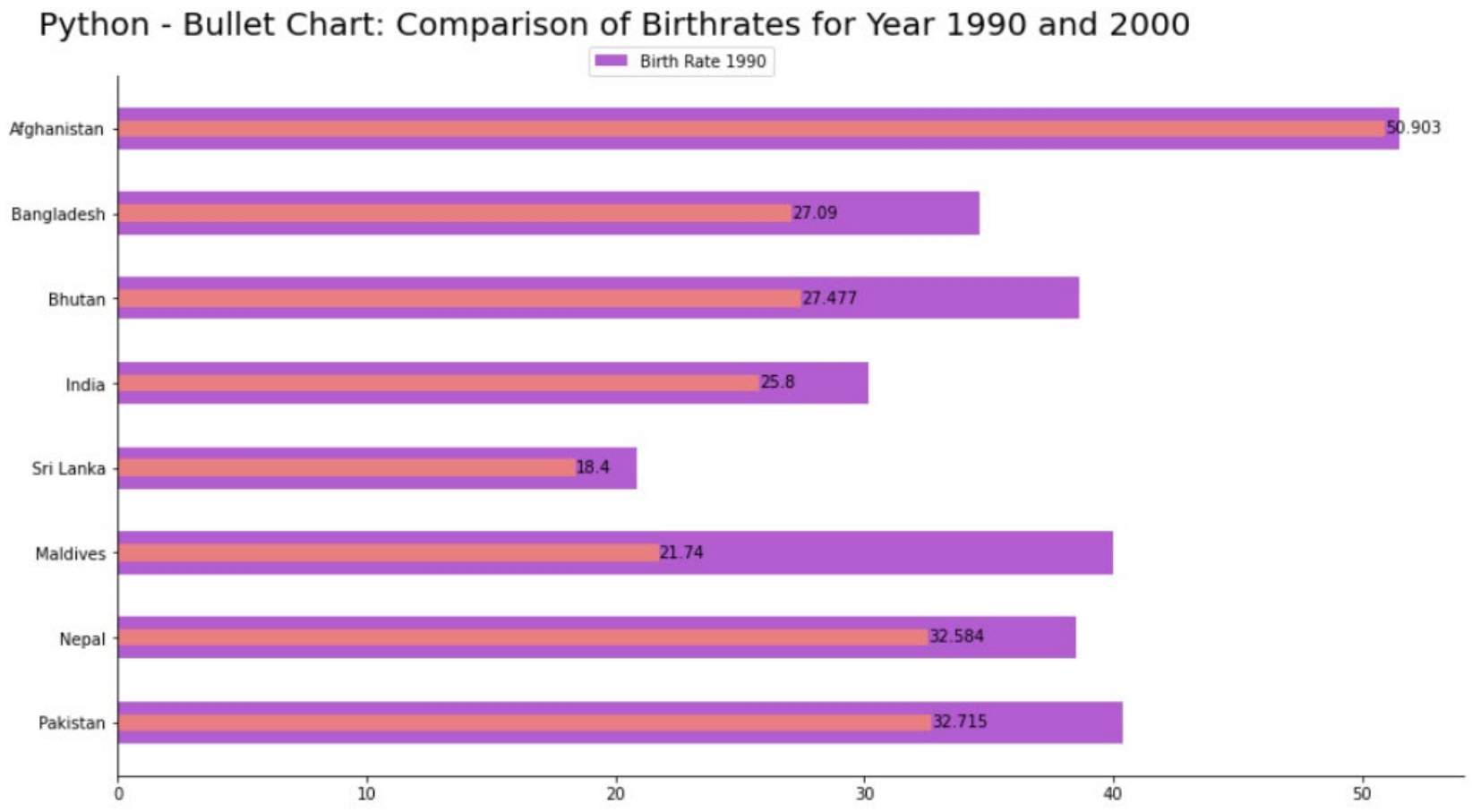
Python – Histogram



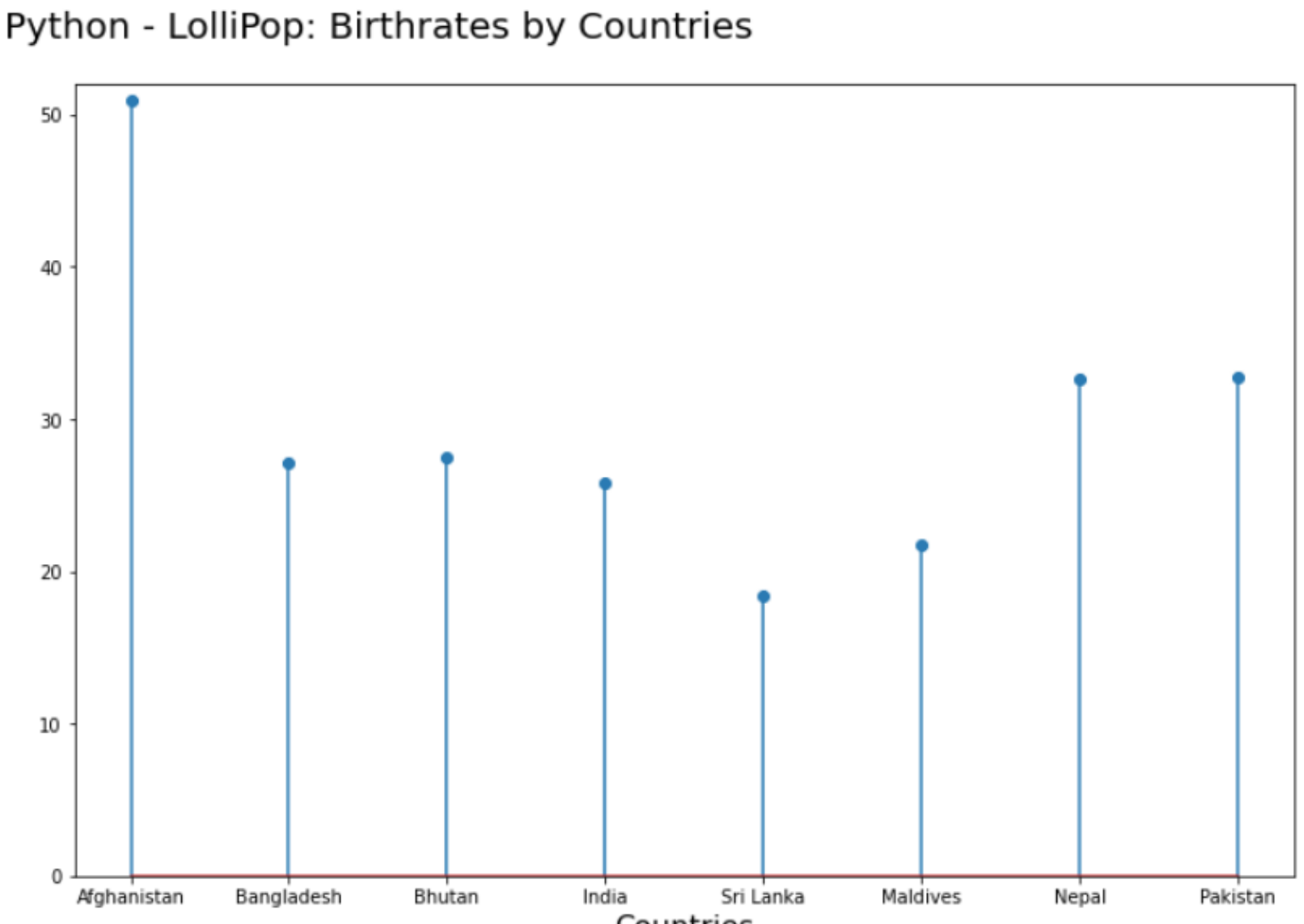
Python – Box Plot



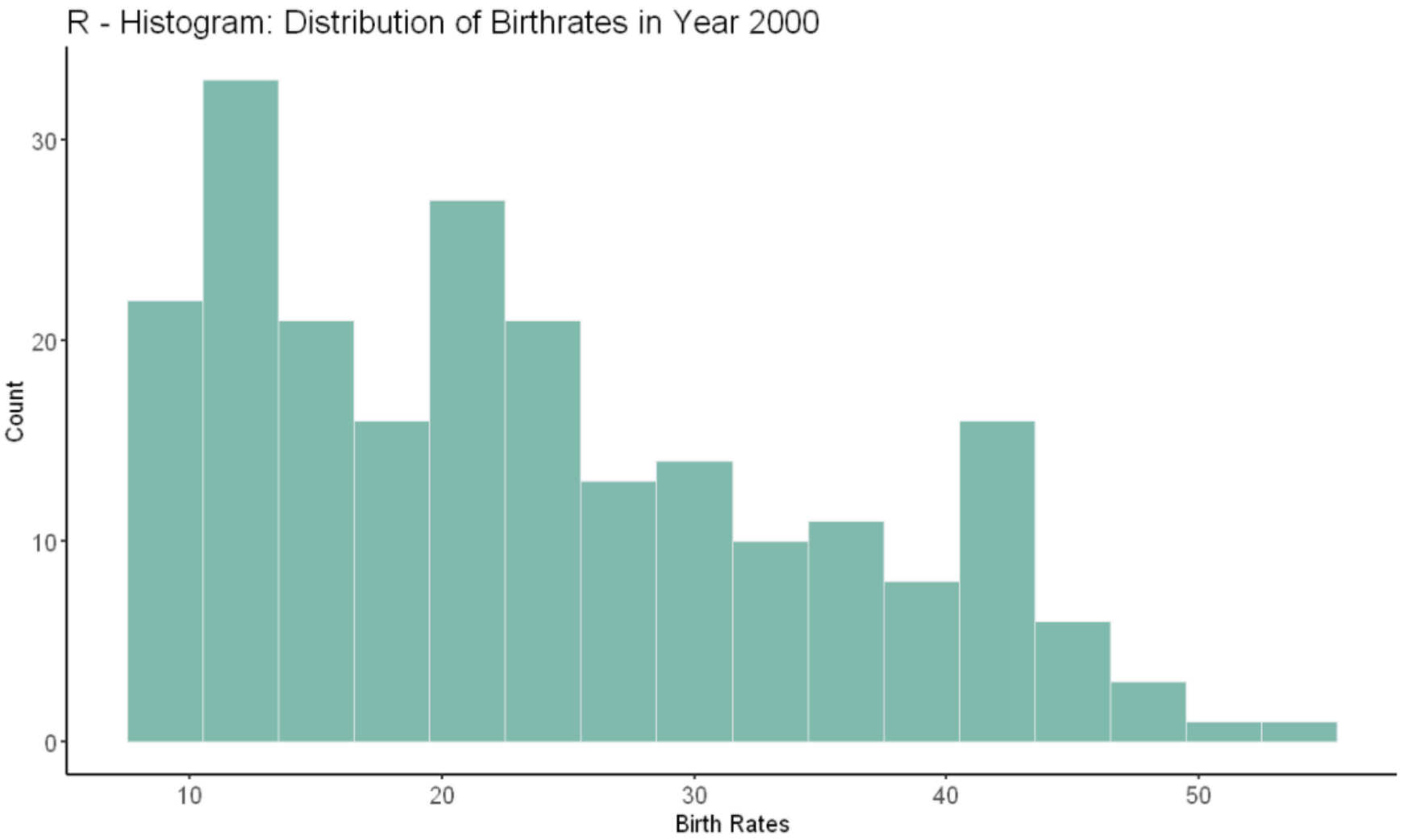
Python – Bullet Chart



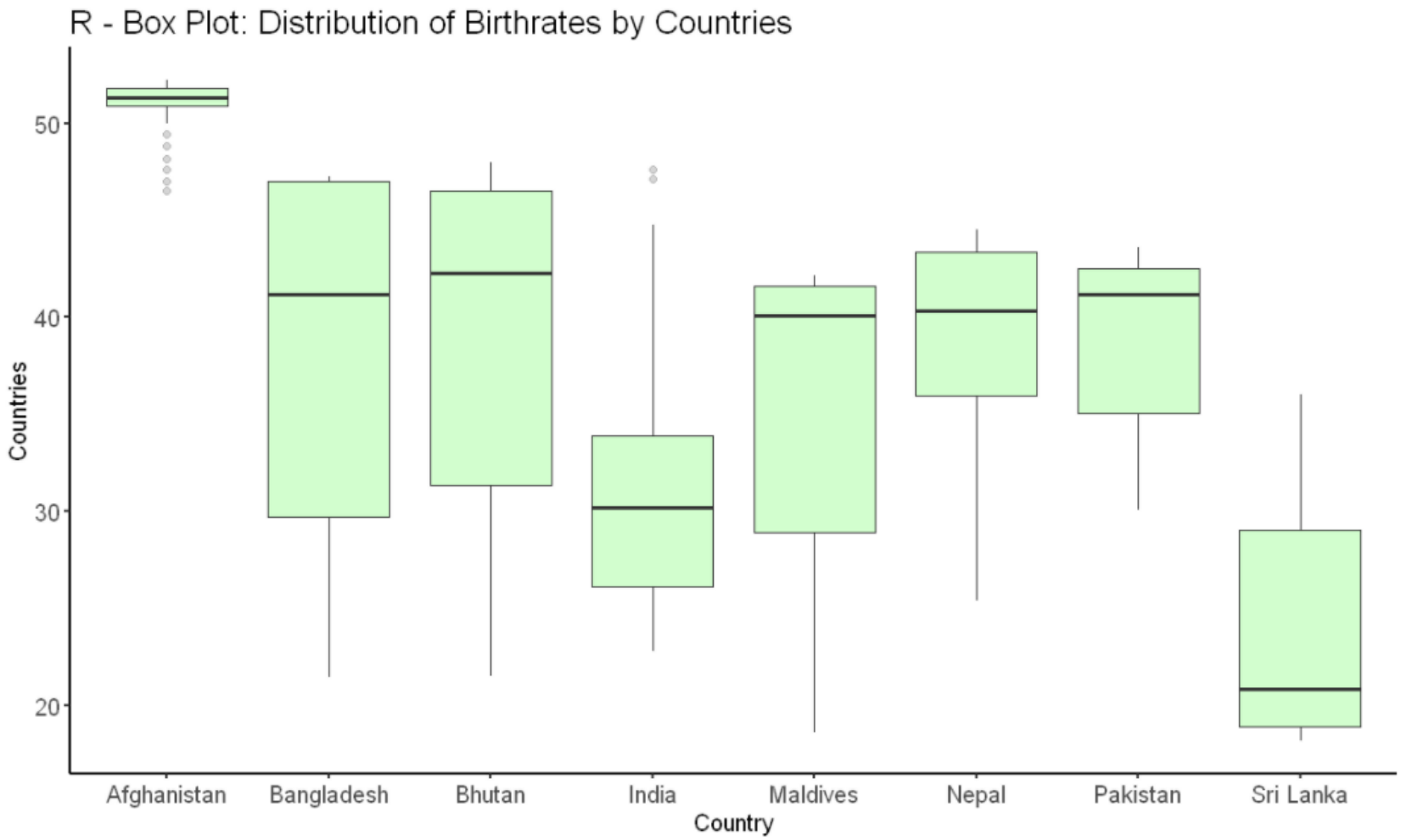
Python – Lolli Pop Chart



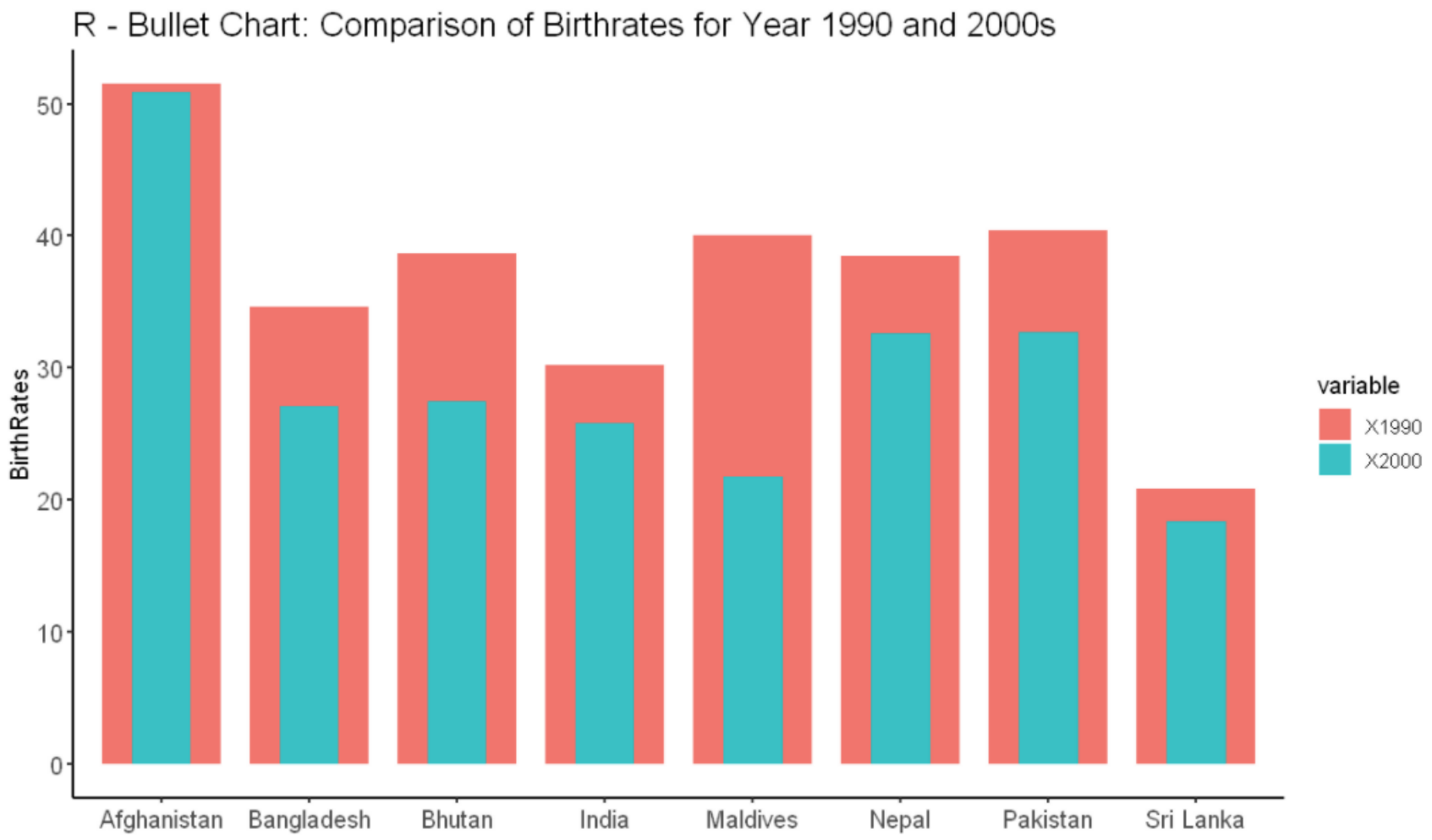
R– Histogram


R - Histogram: Distribution of Birthrates in Year 2000
Count
0
10
20
30
10
20
30
40
50
Birth Rates

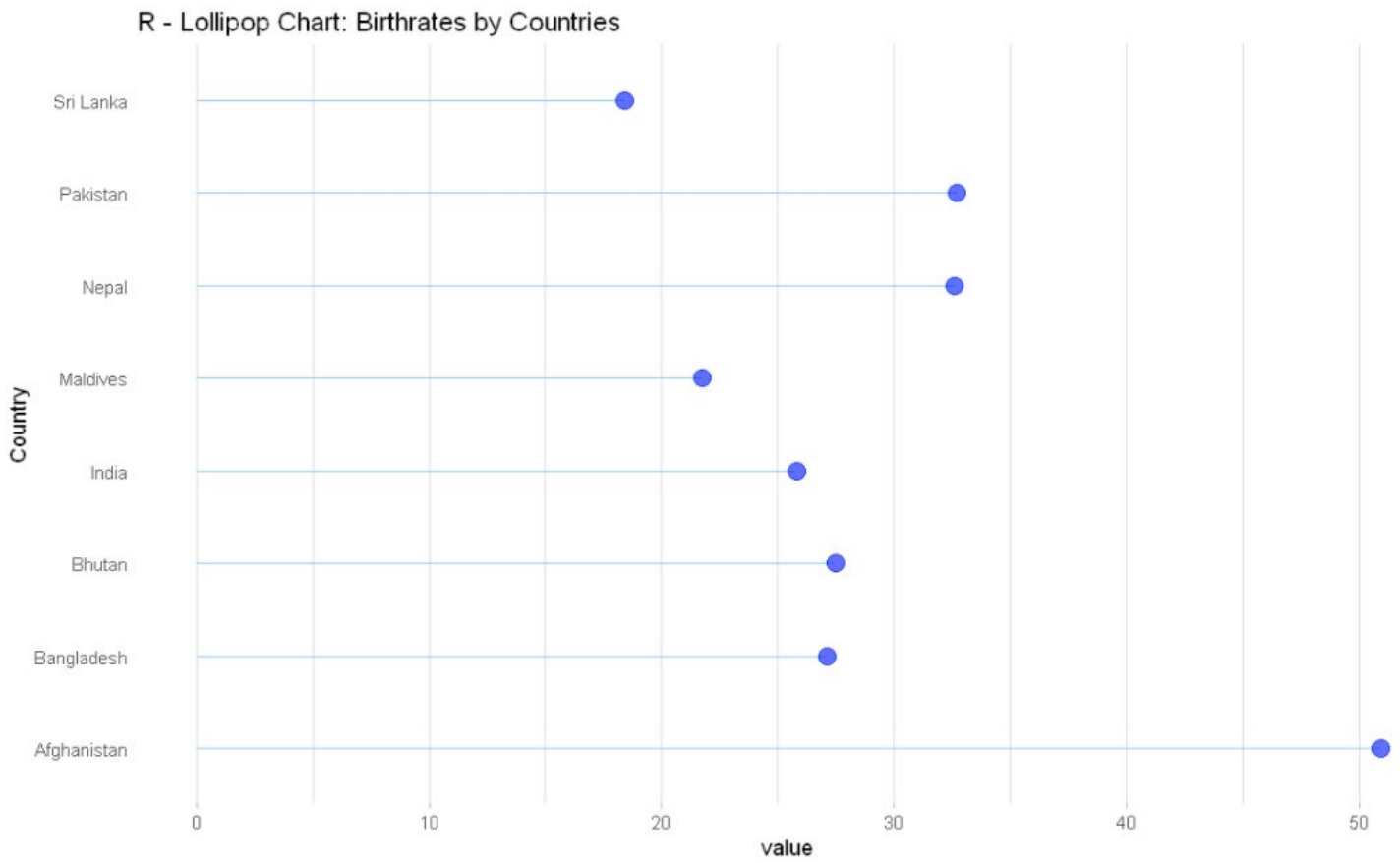
R – Boxplot


R - Box Plot: Distribution of Birthrates by Countries
Countries
50
40
30
20
Afghanistan
Bangladesh
Bhutan
India
Maldives
Nepal
Pakistan
Sri Lanka
Country

R – Bullet Chart



R – LolliPop Chart



Supplemental Files

- Tableau Screenshot
- Python Code
- R Code

Tableau - Book1

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help

Dashboard Layout

Default
Phone

Device Preview

Size

Desktop Browser (1000 x 800)

Sheets

- Histogram
- Boxplot
- Bullet
- Gant

Objects

- Horizontal
- Vertical
- Text
- Image
- Web Page
- Blank
- Navigation
- Download
- Extension
- Ask Data

Tiled Floating

Show dashboard title

Tableau - Histogram: Distribution of Birthrates in Year 2000

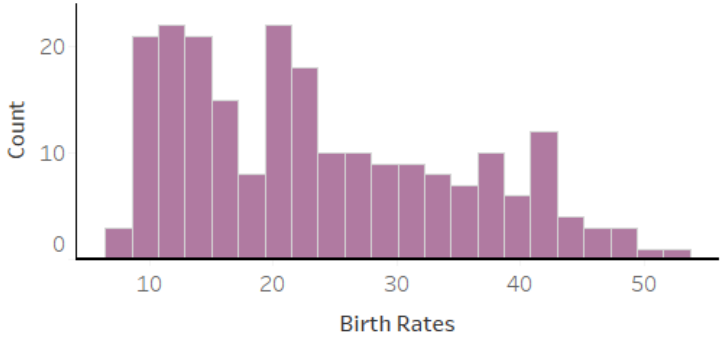


Tableau - Bullet Chart: Comparison of Birthrates of 1990 and 2000 For South Asian Countries

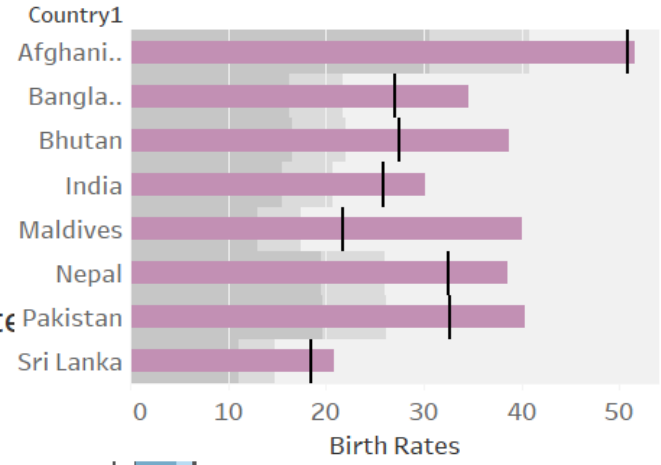


Tableau - Boxplot: Distribution of Birthrate 2000 For South Asian Countries

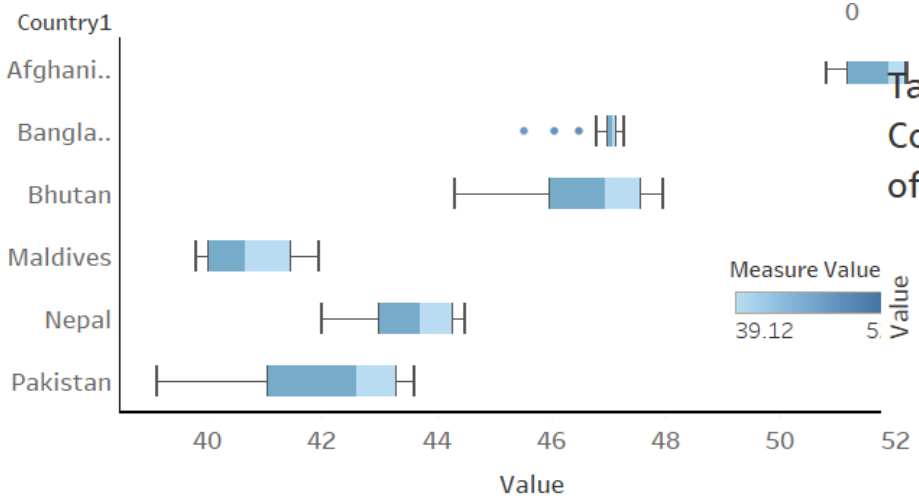
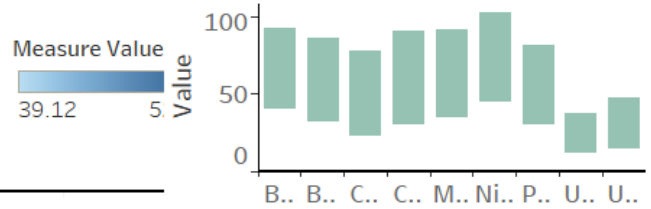


Tableau - Gantt Chart: Comparison of Birthrates of Few American Countries



Data Source | Histogram | Boxplot | Bullet | Gant | Dashboard 1

Python Script

Assignment 6.2: Histogram, Boxplot, Bullet Chart, Parallel Coordinate

DSC640

Taniya Adhikari

In [48]:
```
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
import plotly.graph_objects as go
```

In [49]:
```
birthrates = pd.read_csv('birth-rate.csv')
birthrates.head(3)
```

Out[49]:

	Country	1960	1961	1962	1963	1964	1965	1966	1967	1968	...	1999	2000	2001	2002	2003	2004	200
0	Aruba	36.400	35.179	33.863	32.459	30.994	29.513	28.069	26.721	25.518	...	15.024	14.528	14.041	13.579	13.153	12.772	12.44
1	Afghanistan	52.201	52.206	52.208	52.204	52.192	52.168	52.130	52.076	52.006	...	51.229	50.903	50.486	49.984	49.416	48.803	48.17
2	Angola	54.432	54.394	54.317	54.199	54.040	53.836	53.585	53.296	52.984	...	48.662	48.355	48.005	47.545	46.936	46.184	45.33

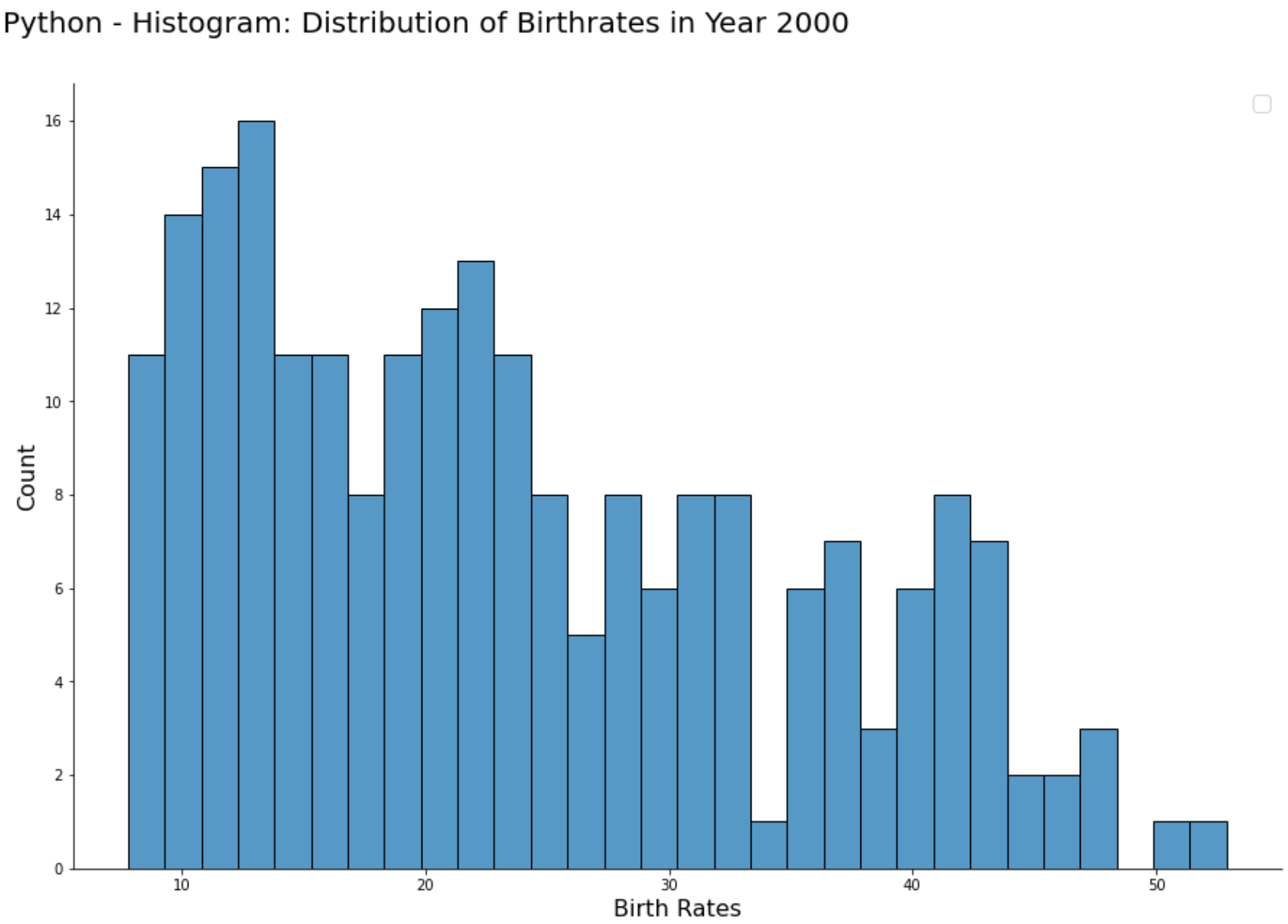
3 rows × 50 columns

Python Histogram

In [50]:
```
plt.rcParams['figure.figsize'] = [15,10]
fig, ax = plt.subplots()
```

```
sns.histplot(data=birthrates, x="2000", bins=30)


plt.suptitle("Python - Histogram: Distribution of Birthrates in Year 2000",
             size=20, x=0.08, y=.95,horizontalalignment='left', verticalalignment='top')

plt.ylabel('Count', size=16)
plt.xlabel('Birth Rates', size=16)

plt.legend(fontsize=15)

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

plt.show()
```

```
No artists with labels found to put in legend.  Note that artists whose label start with an underscore are ignored when
legend() is called with no argument.
```



Python - Box Plot

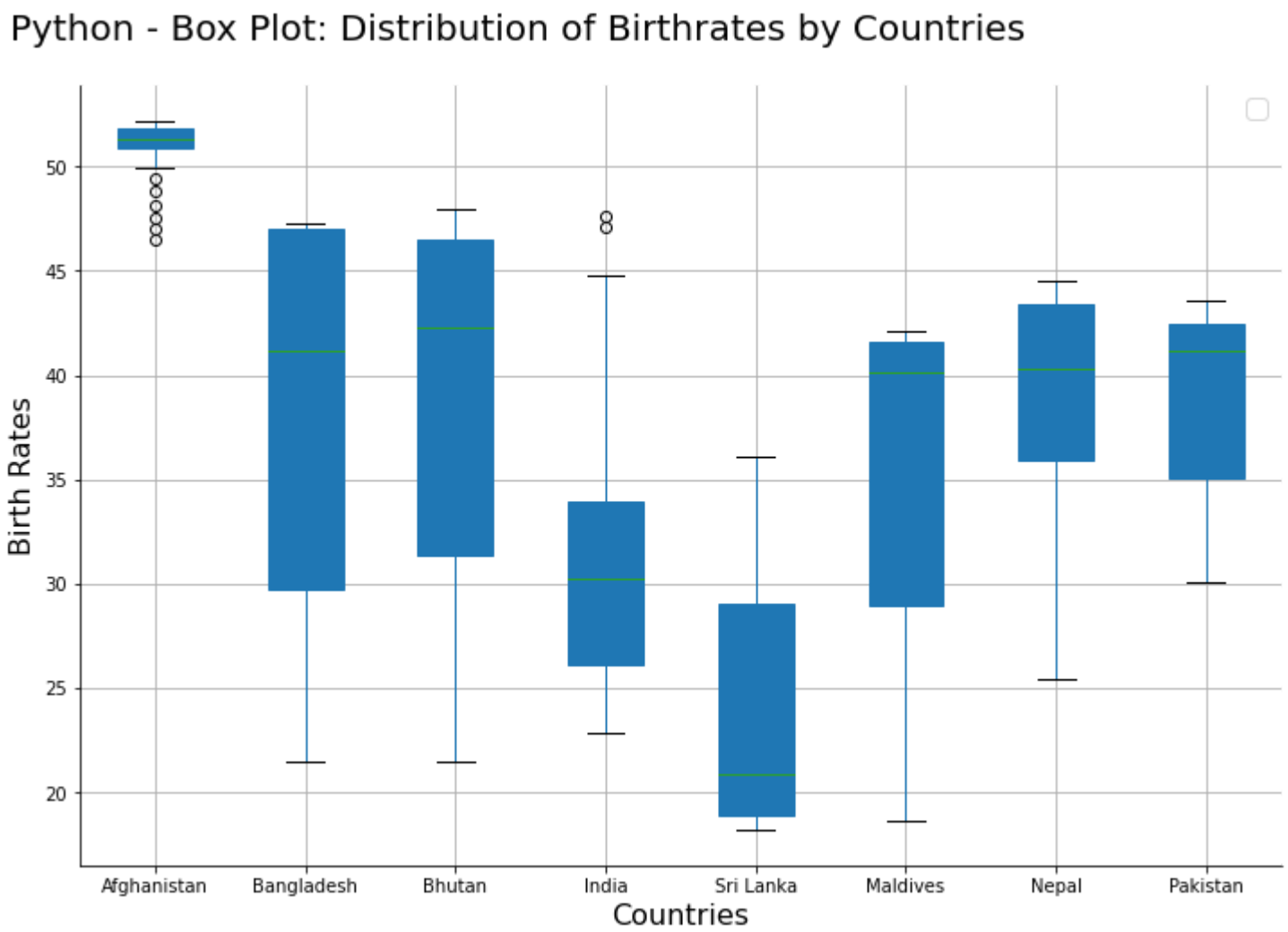
In [51]:
```
southasian = ['India', 'Pakistan', 'Bangladesh', 'Nepal', 'Bhutan', 'Maldives', 'Afghanistan', 'Sri Lanka']
```

In [53]:
```
bd = birthrates.set_index('Country')
```

In [54]:
```
Filter_df  = bd[bd.index.isin(southasian)]
```

In [55]:
```
plt.rcParams['figure.figsize'] = [12,8]

ax = Filter_df.T.boxplot(patch_artist=True)
plt.suptitle("Python - Box Plot: Distribution of Birthrates by Countries",
             size=20, x=0.08, y=.95,horizontalalignment='left', verticalalignment='top')

plt.ylabel('Birth Rates', size=16)
plt.xlabel('Countries', size=16)

plt.legend(fontsize=15)

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

```
No artists with labels found to put in legend.  Note that artists whose label start with an underscore are ignored when
legend() is called with no argument.
```



Python - Bullet Chart

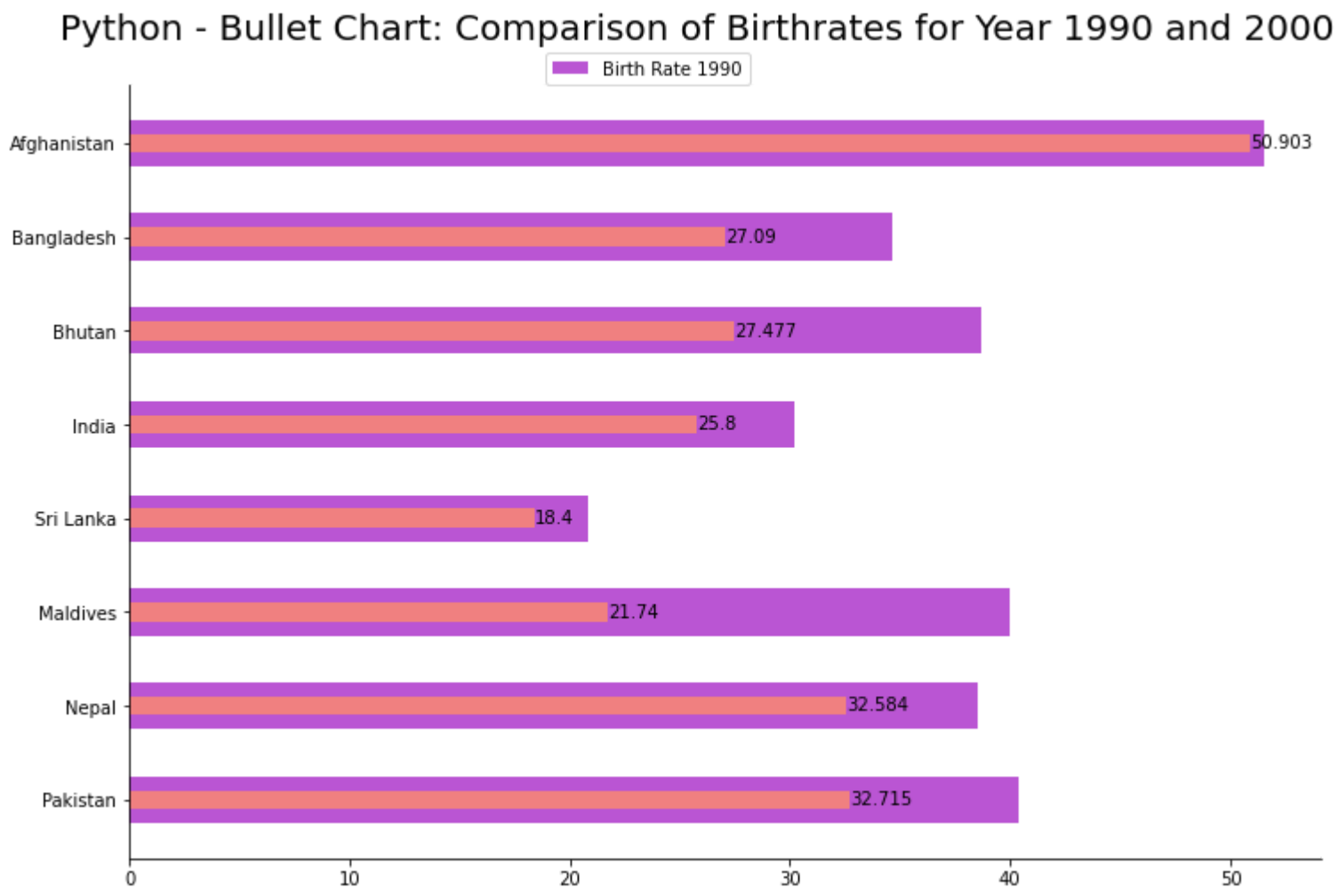
In [74]:

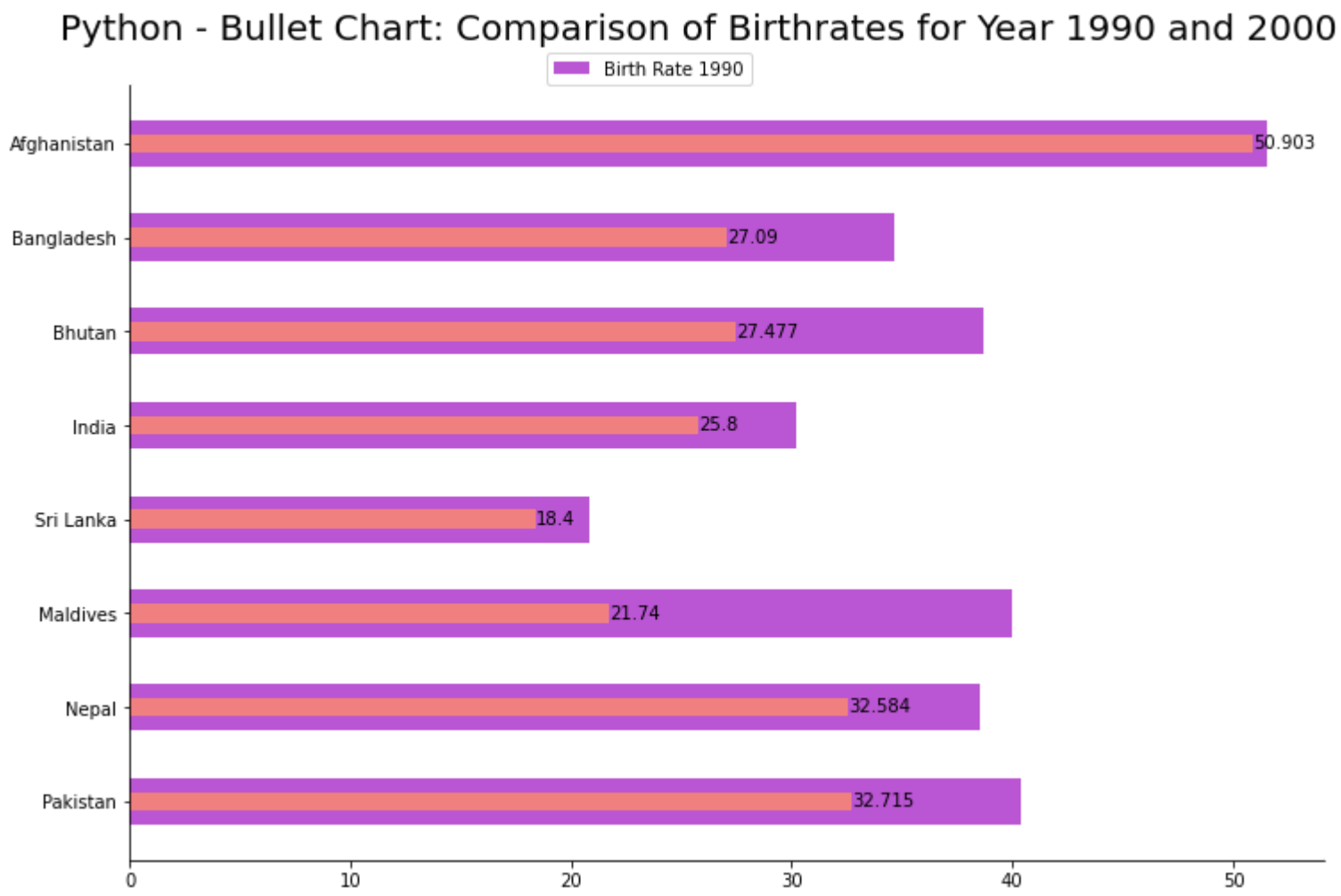
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

# Function: bullet_chart, creates a horizontal bar graph representing a bullet chart
# Inputs: (1) dataframe with 4 rows - the name of the column ("business"), "base goal" or the quota, "stretch goal" or
# Output: matplotlib image representing a bullet chart
def bullet_chart(df,color_code=False):
    y_pos = np.arange(len(df.index))
```

```
  #assign coloring
    df["col"]="indigo"
    if (color_code==True):
        for i in y_pos:
            if(df["2000"][i]>=df["1990"][i]):
                df["col"][i]="gold"
            else:
                df["col"][i]="lightcoral"

    #Initialize plot
    fig, ax = plt.subplots()
    ax.barh(y_pos, df["1990"], height=0.5, align='center', color='mediumorchid', label = "Birth Rate 1990")
    ax.barh(y_pos, df["2000"], height=0.2, align='center',color=df["col"])
    ax.set_yticklabels(df.index)
    ax.set_yticks(y_pos)
    ax.invert_yaxis()

  #add data labels
    for i in y_pos:
        ax.text(df["2000"][i], i+0.05, df["2000"][i])

      #add legend and format borders
    plt.legend(loc=(0.35,1.0))
    plt.suptitle("Python - Bullet Chart: Comparison of Birthrates for Year 1990 and 2000",
             size=20, x=0.08, y=.95,horizontalalignment='left', verticalalignment='top')

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

    return fig

# Use Sisense for Cloud Data Teams to visualize a dataframe or an image by passing data to periscope.output()
bullet_chart(bullet, color_code=True)
```

```
C:\Users\bibek\AppData\Local\Temp\ipykernel_4264\87858703.py:12: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a
-view-versus-a-copy
  df["col"]="Blue"
C:\Users\bibek\AppData\Local\Temp\ipykernel_4264\87858703.py:18: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a
-view-versus-a-copy
  df["col"][i]="lightcoral"
C:\Users\bibek\AppData\Local\Temp\ipykernel_4264\87858703.py:24: UserWarning: FixedFormatter should only be used togeth
er with FixedLocator
  ax.set_yticklabels(df.index)
```

Out[74]:


Python - Bullet Chart: Comparison of Birthrates for Year 1990 and 2000
Birth Rate 1990
Afghanistan
Bangladesh
Bhutan
India
Sri Lanka
Maldives
Nepal
Pakistan
50.903
27.09
27.477
25.8
18.4
21.74
32.584
32.715
0
10
20
30
40
50

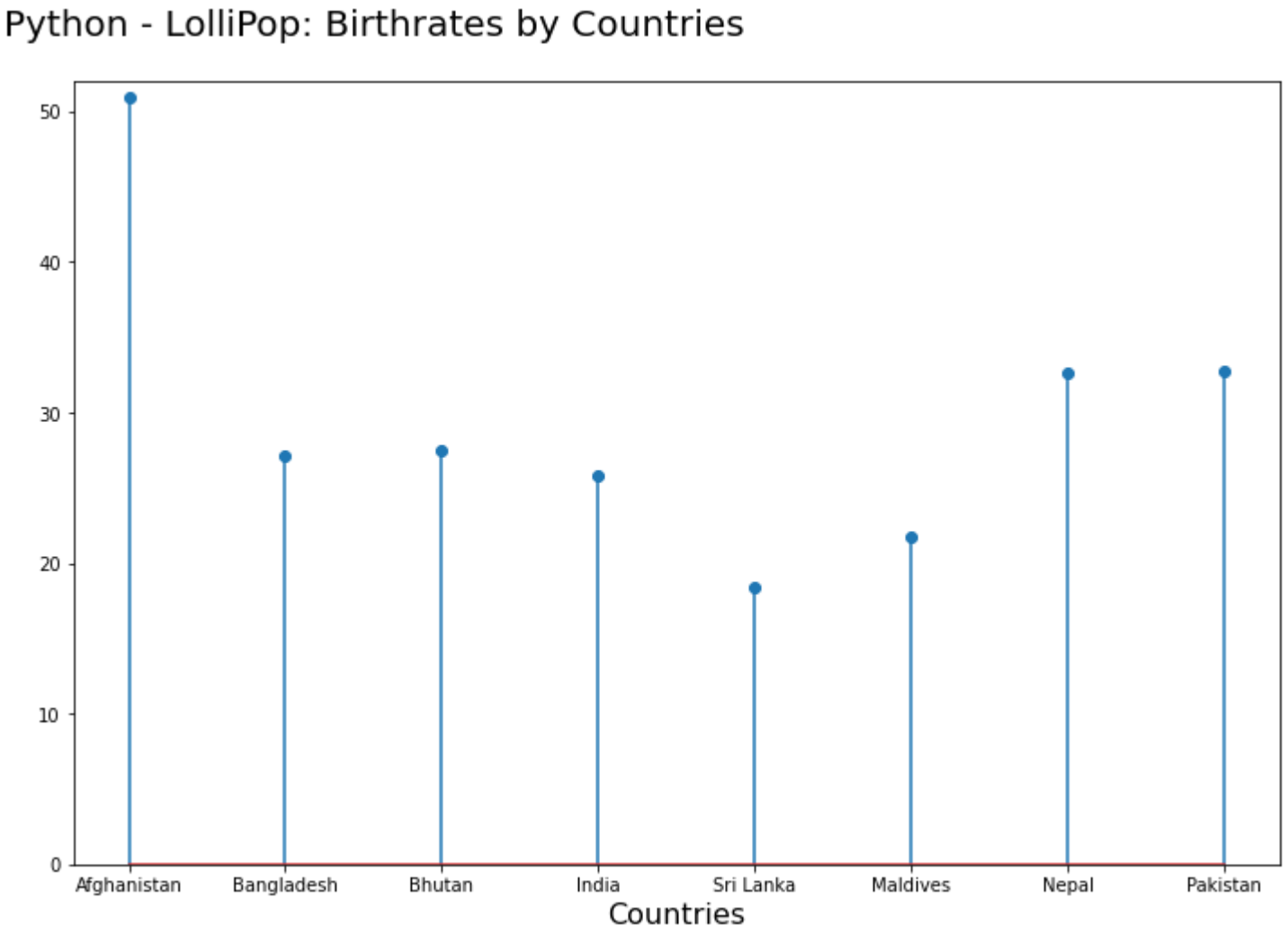


Python - Lollipop Plot:

In [75]:

```
plt.rcParams['figure.figsize'] = [12,8]

# stem function
ax = plt.stem(bullet.index, bullet['2000'])
plt.ylim(0, 52)
plt.suptitle("Python - LolliPop: Birthrates by Countries",
             size=20, x=0.08, y=.95,horizontalalignment='left', verticalalignment='top')

plt.xlabel('Countries', size=16)
```

Out[75]:
```
Text(0.5, 0, 'Countries')
```



In [ ]:

R Script

Assignment 6.2: Histogram, Boxplot, Bullet Chart, Parallel Coordinate

DSC640

Taniya Adhikari

In [35]:
```
library(ggplot2)
library(readxl)
library(scales)
library(plyr)
library(dplyr)
library(ggrepel)
library(reshape2)
```

In [7]:
```
birthrates <- read.csv('birth-rate.csv')
head(birthrates)
```

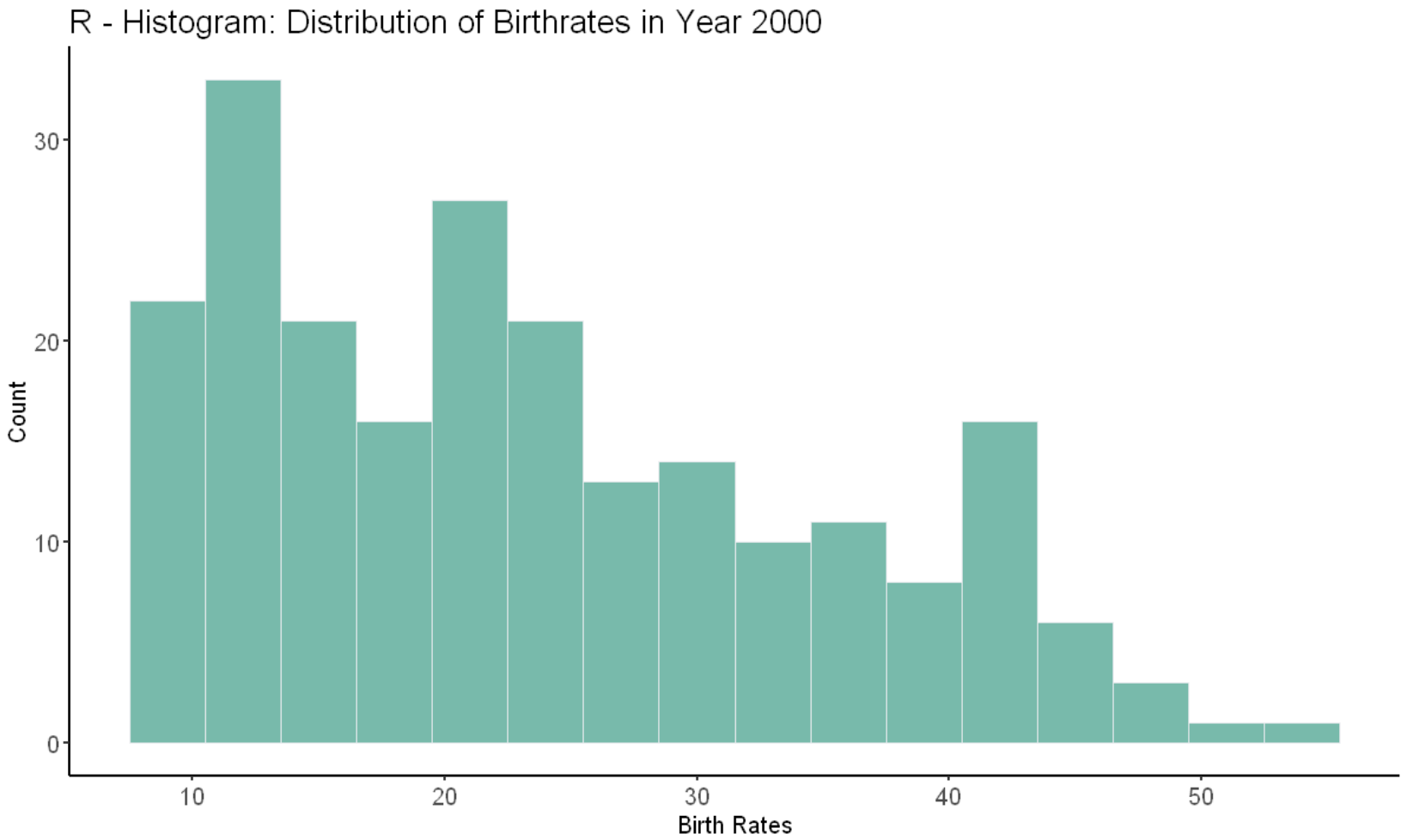
Country	X1960	X1961	X1962	X1963	X1964	X1965	X1966	X1967	X1968	...	X1999	X2000	X2001	X2002	
Aruba	36.40000	35.179	33.863	32.459	30.994	29.51300	28.069	26.721	25.518	...	15.02400	14.5280	14.04100	13.57900	13
Afghanistan	52.20100	52.206	52.208	52.204	52.192	52.16800	52.130	52.076	52.006	...	51.22900	50.9030	50.48600	49.98400	49
Angola	54.43200	54.394	54.317	54.199	54.040	53.83600	53.585	53.296	52.984	...	48.66200	48.3550	48.00500	47.54500	46
Albania	40.88600	40.312	39.604	38.792	37.913	37.00800	36.112	35.245	34.421	...	17.71300	16.8500	16.08100	15.44400	14
Netherlands Antilles	32.32100	30.987	29.618	28.229	26.849	25.51800	24.280	23.173	22.230	...	15.80900	15.4120	15.09600	14.82400	14
Arab World	47.61122	NA	NA	NA	NA	46.57288	NA	NA	NA	...	29.07314	28.6828	28.30524	27.96244	27

R - Histogram

In [21]:

```
options(repr.plot.width =10, repr.plot.height =6)

# plot
p <- birthrates %>%
  ggplot( aes(x=X2000)) +
    geom_histogram( binwidth=3, fill="#69b3a2", color="#e9ecef", alpha=0.9) +
    ggtitle("Bin size = 30")
p + 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))+
ylab("Count") +
xlab("Birth Rates") +
  ggtitle("R - Histogram: Distribution of Birthrates in Year 2000")

```

```
Warning message:
"Removed 11 rows containing non-finite values (stat_bin)."
```



R - Bubble Chart

In [33]:

```
1 dt <-birthrates %>%
2             filter(Country %in% c('India', 'Pakistan', 'Bangladesh', 'Nepal', 'Bhutan', 'Maldives', 'Afg
```

In [40]:

```
dt2 <- melt(dt)
head(dt2)
```

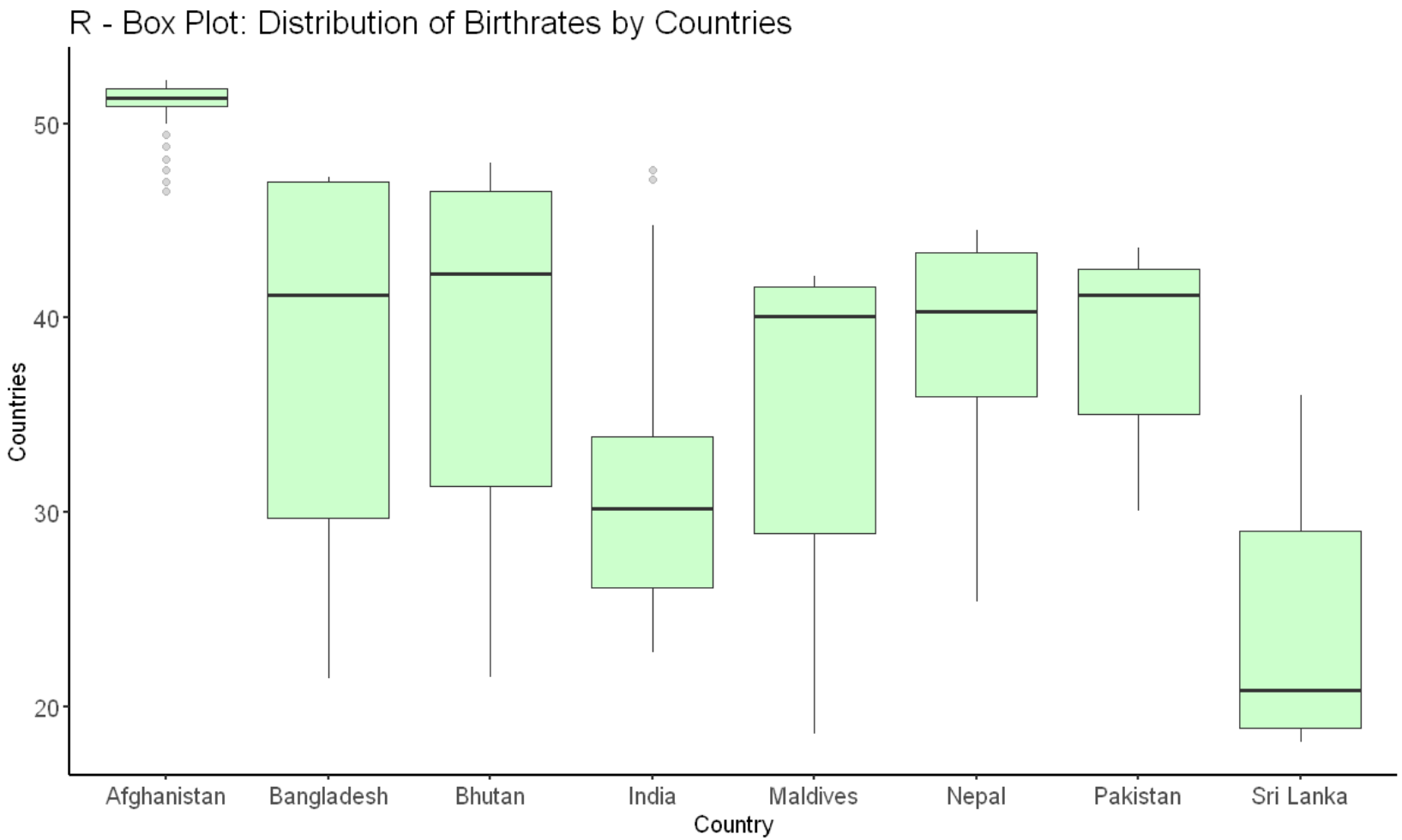
```
Using Country as id variables
```

Country	variable	value
Afghanistan	X1960	52.201
Bangladesh	X1960	47.258
Bhutan	X1960	47.945
India	X1960	47.580
Sri Lanka	X1960	36.046
Maldives	X1960	41.741

In [51]:

```
ggplot(dt2, aes(x=Country, y=value)) +
    geom_boxplot(fill='green', alpha=.2) +
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))+
ylab("BirthRates") +
ylab("Countries") +
ggtitle("R - Box Plot: Distribution of Birthrates by Countries")
```

```
Warning message:
"Removed 36 rows containing non-finite values (stat_boxplot)."
```



R - Density Plot

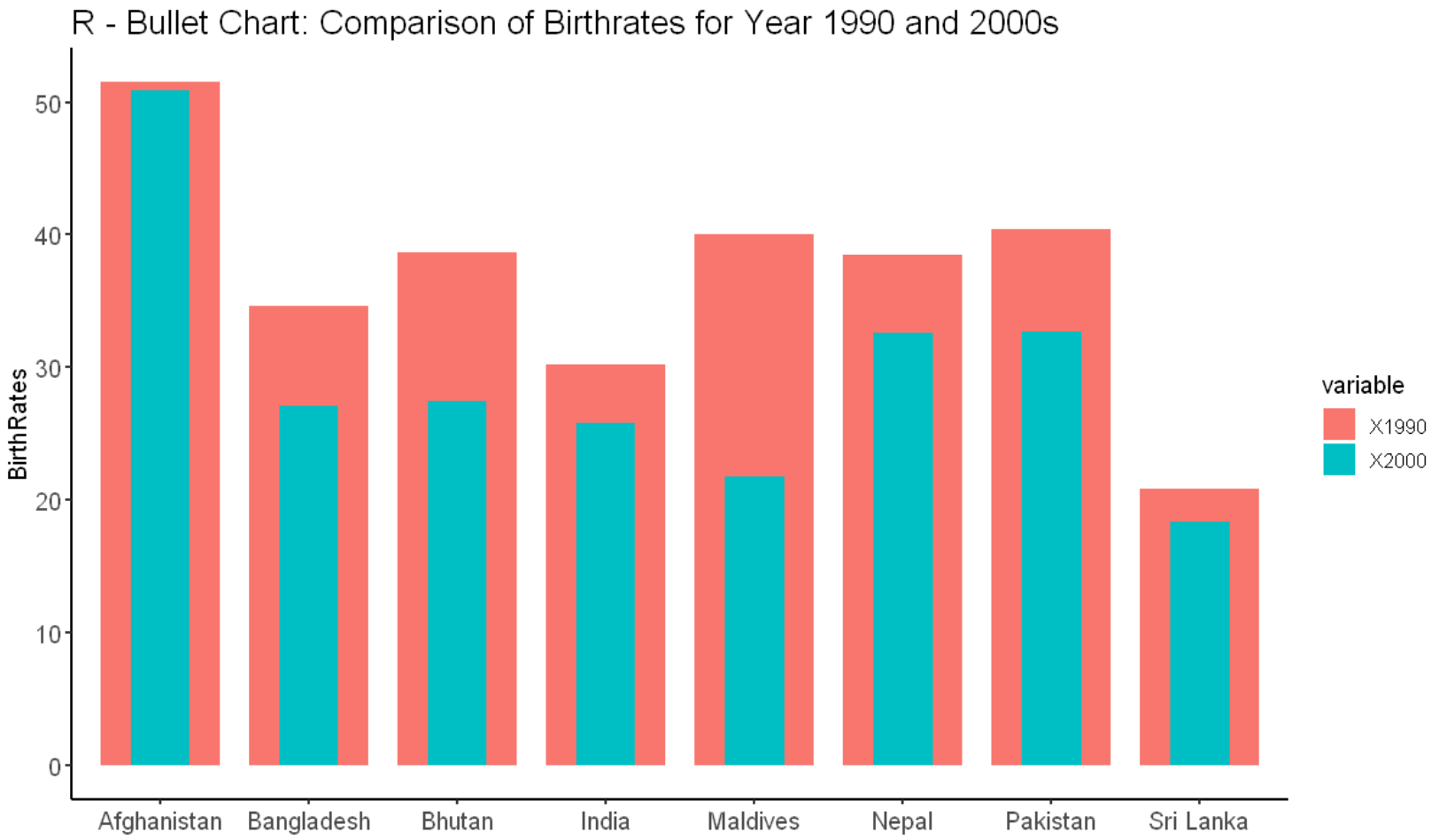
In [54]:
```
dt3 <-dt2 %>%
            filter(variable %in% c('X1990', 'X2000'))
```

In [59]:
```
dt4<-transform(dt3,width=ifelse(variable=="X1990",.8,.4))
```

In [68]:

```
gg <- ggplot(dt4)
gg <- gg + geom_bar(aes(Country, value, fill=variable), width=dt4$width, stat="identity")
gg <- gg + xlab("") + ylab("")+
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))+
ylab("BirthRates") +
ggtitle("R - Bullet Chart: Comparison of Birthrates for Year 1990 and 2000s")


print(gg)

```

```
Warning message:
"position_stack requires non-overlapping x intervals"
```



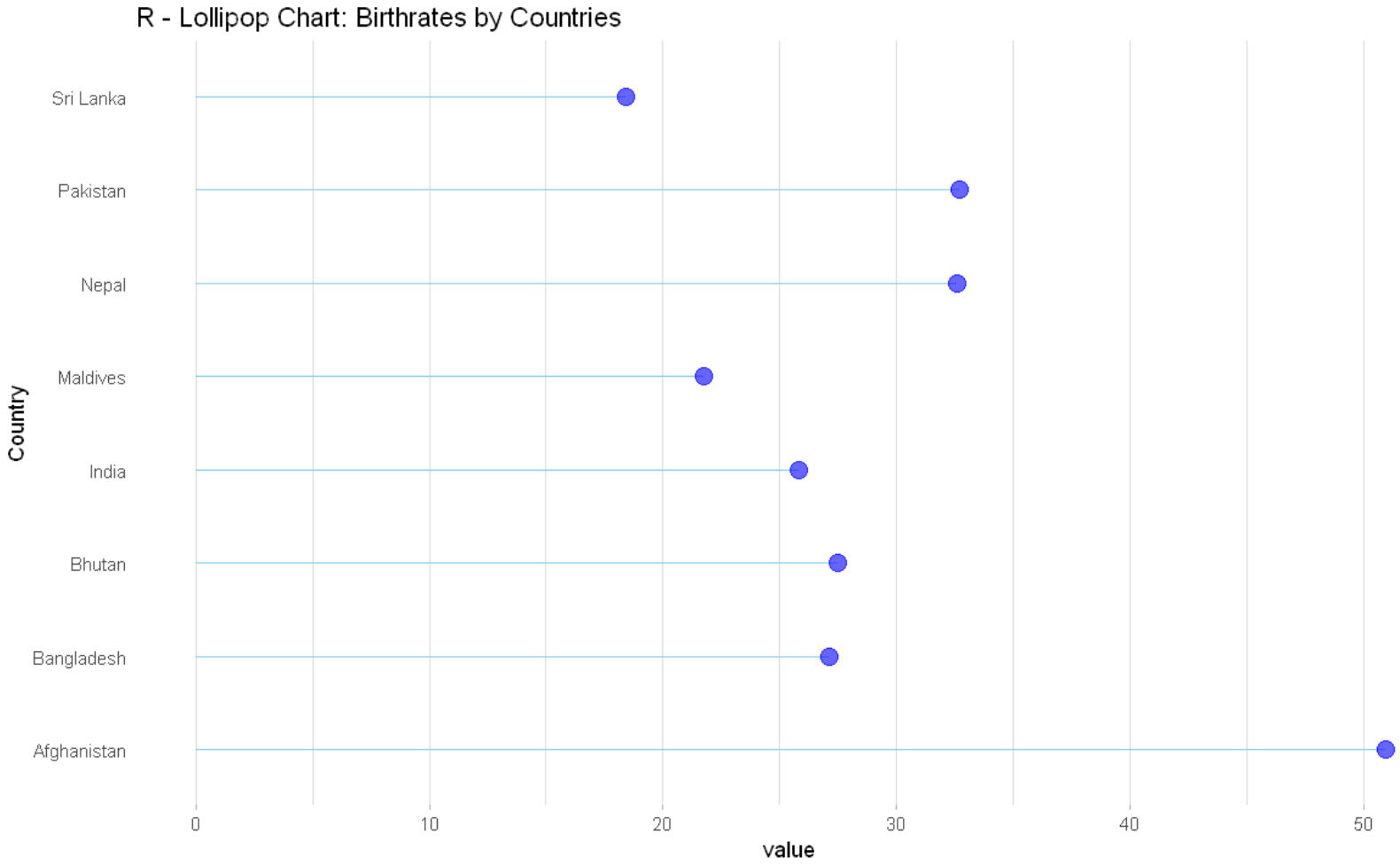
R - Lolliop Chart

In [65]:

```
1 dt5 <-dt3 %>%
2             filter(variable %in% c('X2000'))
```

In [69]:

```
# Horizontal version
ggplot(dt5, aes(Country, y=value)) +
  geom_segment( aes(x=Country, xend=Country, y=0, yend=value), color="skyblue") +
  geom_point( color="blue", size=4, alpha=0.6) +
  theme_light() +
  coord_flip() +
  theme(
    panel.grid.major.y = element_blank(),
    panel.border = element_blank(),
    axis.ticks.y = element_blank()
  )+
ggtitle("R - Lollipop Chart: Birthrates by Countries")
```



In [ ]:

```
1
```