**Ryan Long**
**DSC640**
**2.2 Exercises: Charts**

**Datasets used:**
world-population.xlsm

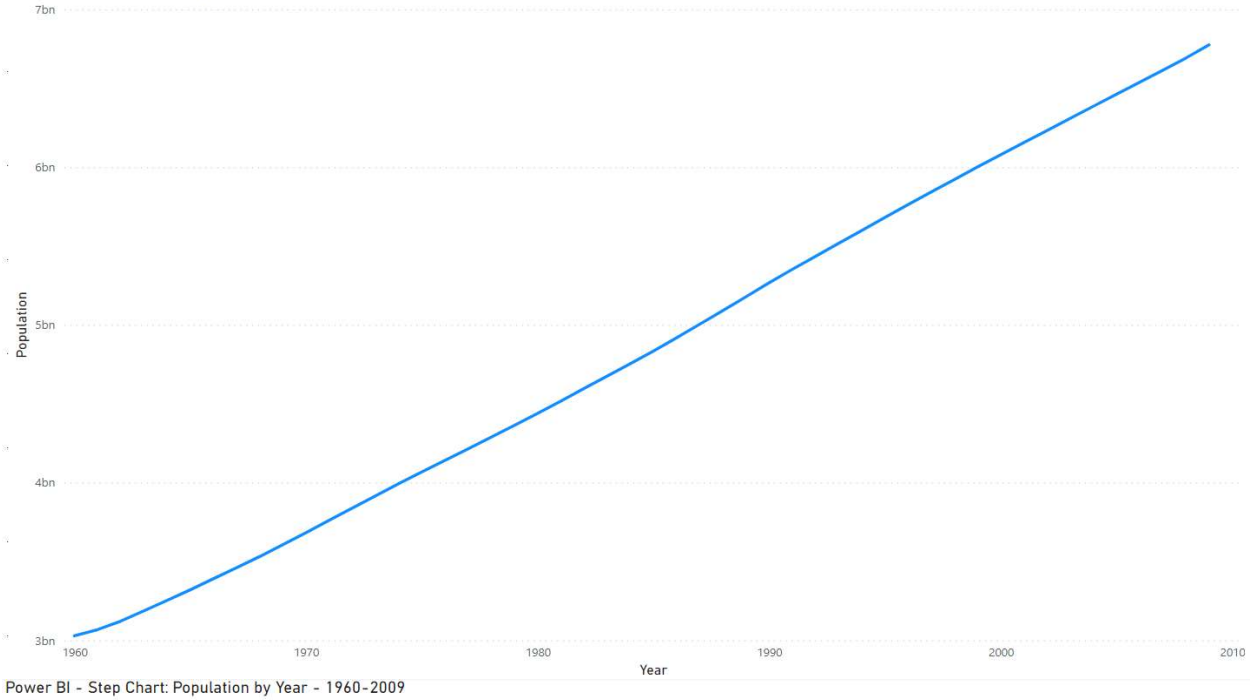
**Summary**
The charts were straight forward this week. The given dataset required few adjustments prior to coding the charts. I did create a column to display the population in billions to simply the chart. We've had discussion in the class Teams group the last few weeks on starting the y-axis at 0. Initially, I agreed it should always start at zero, but now after this exercise I feel it should be situation dependent. When comparing two things, it likely should start at zero. But in this situation for this dataset, I think it make sense to start nearer the first datapoint. A reader wouldn't assume the world started at 1960 with approximately 3 billion people. I'm glad this exercise caused me to rethink my initial opinions on one of our discussion topics.

As far as the type of chart for this data, I prefer the step chart. I think it shows greater variation and confirms with the reader the data isn't a mistake. With the simple line chart, aside from a slight undulation in the line, a reader may think there is something off with the data.

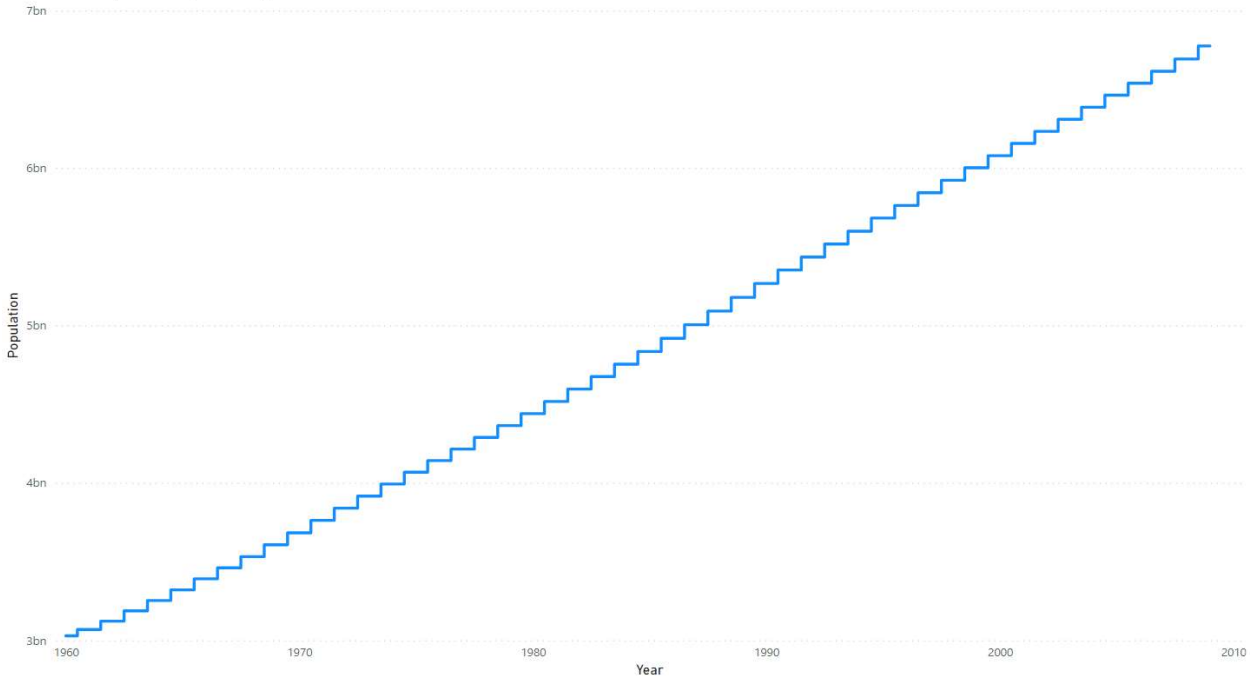
**The following pages contain:**

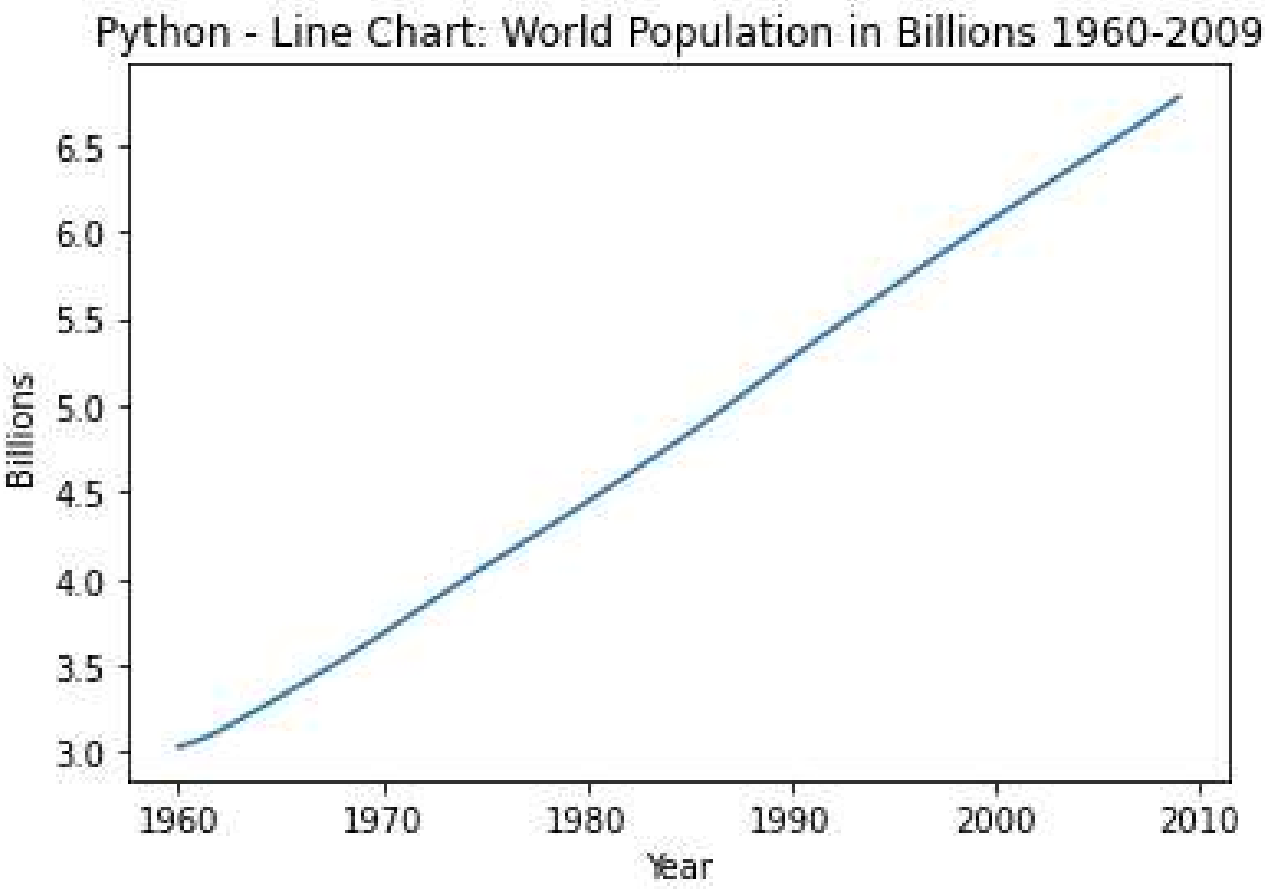

**Appendix**
Code support for both Python and R notebooks

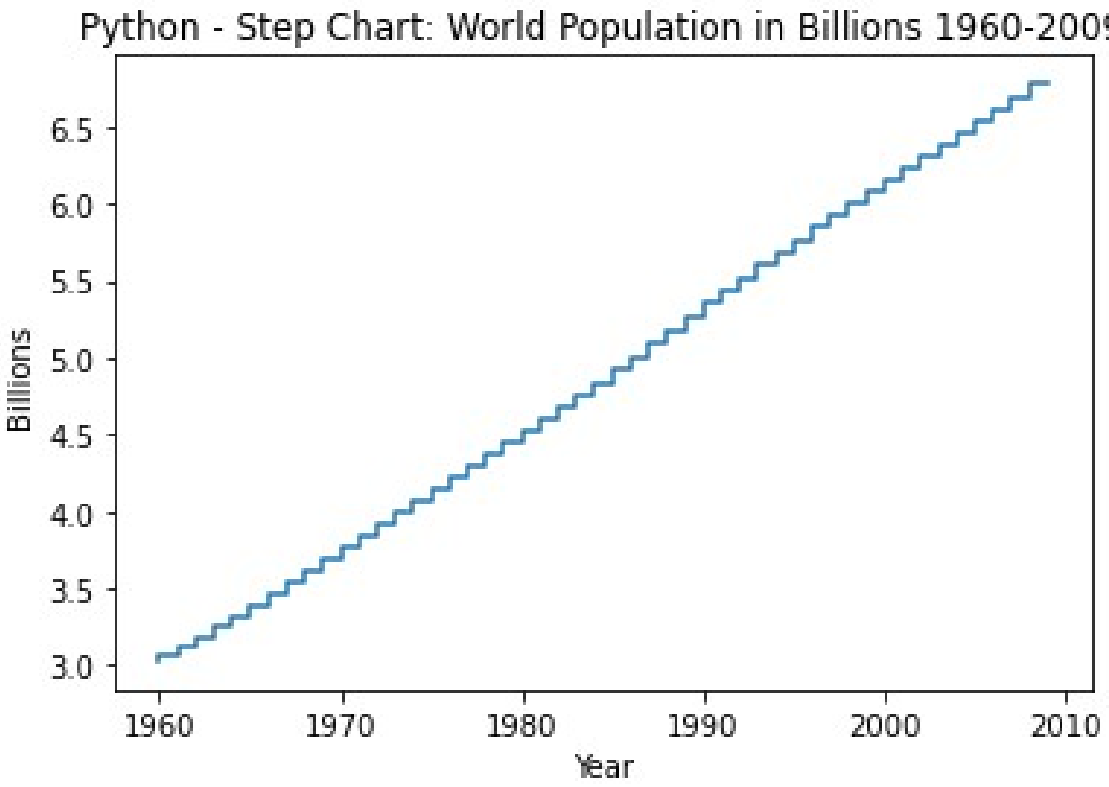
Power BI - Line Chart: Population by Year - 1960-2009



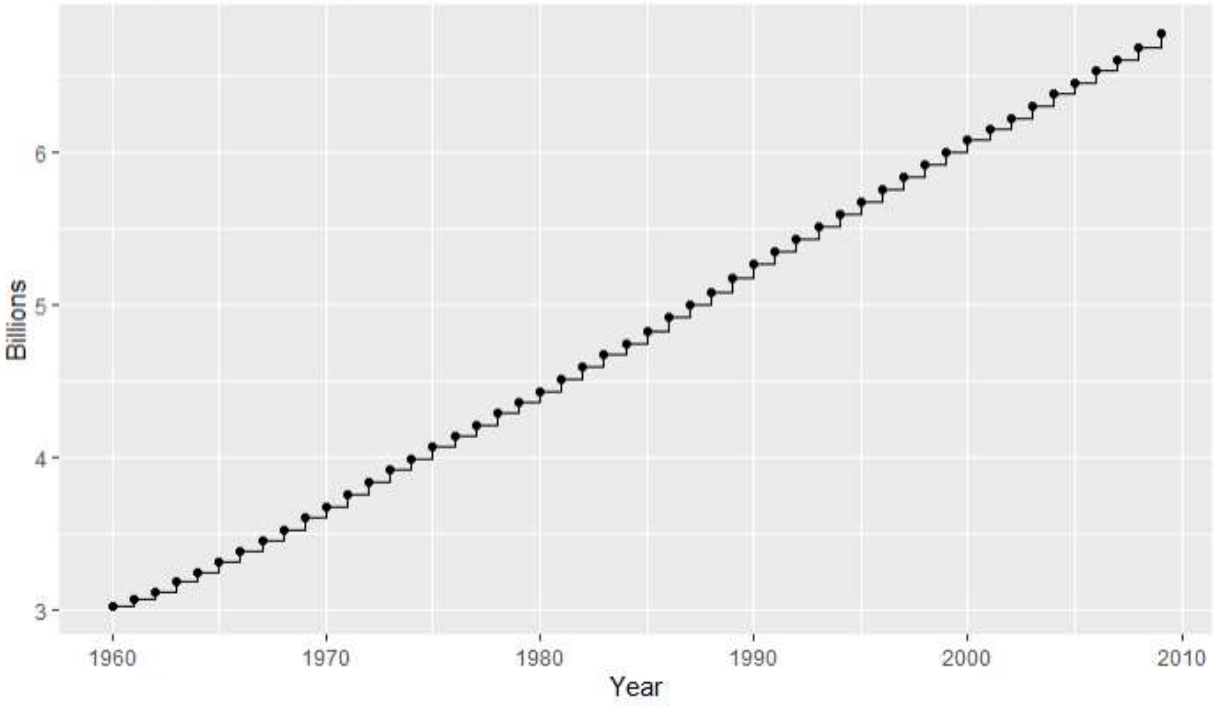

Power BI - Step Chart: Population by Year - 1960-2009



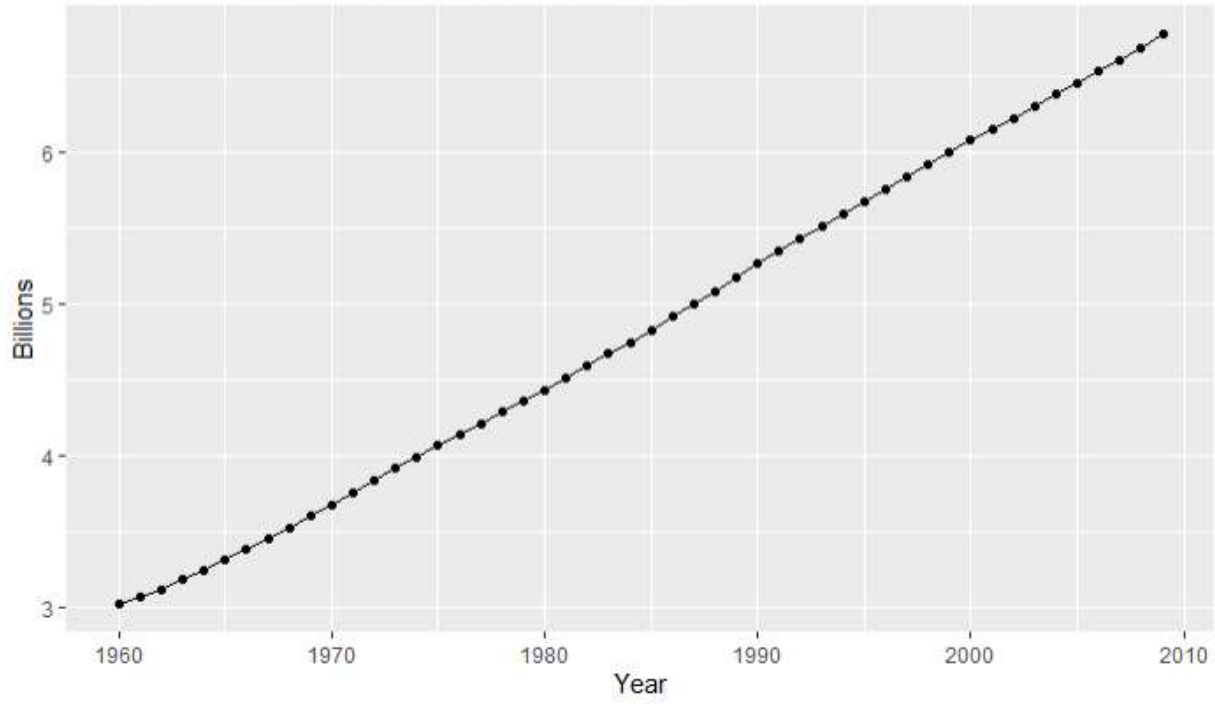

Python - Line Chart: World Population in Billions 1960-2009
Billions
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
1960
1970
1980
1990
2000
2010
Year


Python - Step Chart: World Population in Billions 1960-2009
Billions
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
1960
1970
1980
1990
2000
2010
Year

R: Step Chart - World Population - 1960-2009




R: Line Chart - World Population - 1960-2009



# APPENDIX

In [1]:
```
#load libraries
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
#import data as dataframe
data = pd.read_excel('world-population.xlsm')
```

In [3]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Year        50 non-null     int64
 1   Population  50 non-null     int64
dtypes: int64(2)
memory usage: 928.0 bytes
```

In [4]:
```
data.head()
```

Out[4]:

| | Year | Population |
|---|---|---|
| **0** | 1960 | 3028654024 |
| **1** | 1961 | 3068356747 |
| **2** | 1962 | 3121963107 |
| **3** | 1963 | 3187471383 |
| **4** | 1964 | 3253112403 |

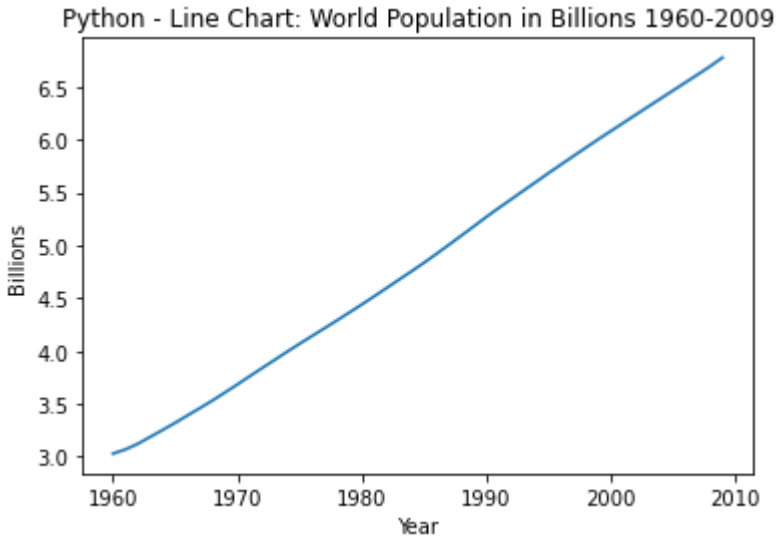
In [5]:
```
#create column for billions
data['PopinB'] = data['Population'] / 1000000000
```

# Line Chart

In [6]:
```
x = data['Year']
y = data['PopinB']

plt.plot(x,y)
plt.xlabel("Year")  # X-axis label
plt.ylabel("Billions")  # Y-axis label
plt.title("Python - Line Chart: World Population in Billions 1960-2009")  # title
```

Out[6]:
```
Text(0.5, 1.0, 'Python - Line Chart: World Population in Billions 1960-2009')
```

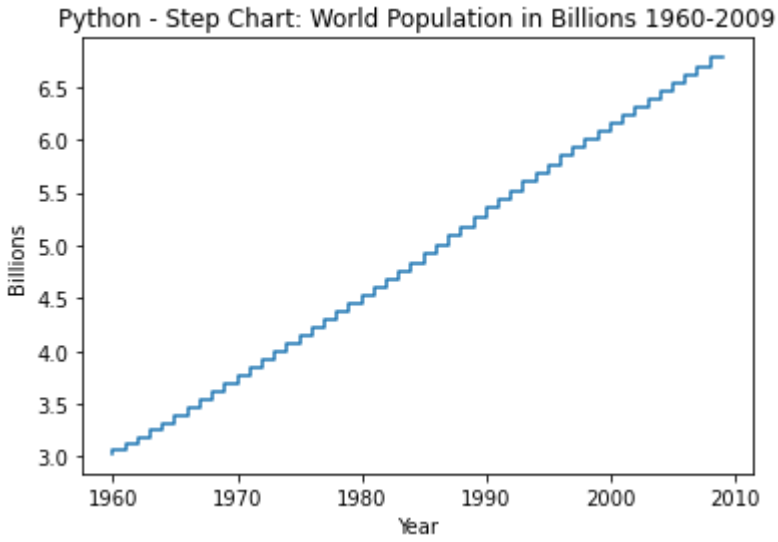



# Step Chart

In [7]:
```
x = data['Year']
y = data['PopinB']

plt.step(x,y)
plt.xlabel("Year")  # X-axis label
plt.ylabel("Billions")  # Y-axis label
plt.title("Python - Step Chart: World Population in Billions 1960-2009")  # title
```

Out[7]:
```
Text(0.5, 1.0, 'Python - Step Chart: World Population in Billions 1960-2009')
```



# Week 3 & 4

Code

Hide

```
#load libraries
library(ggplot2)
library(readxl)
```

Hide

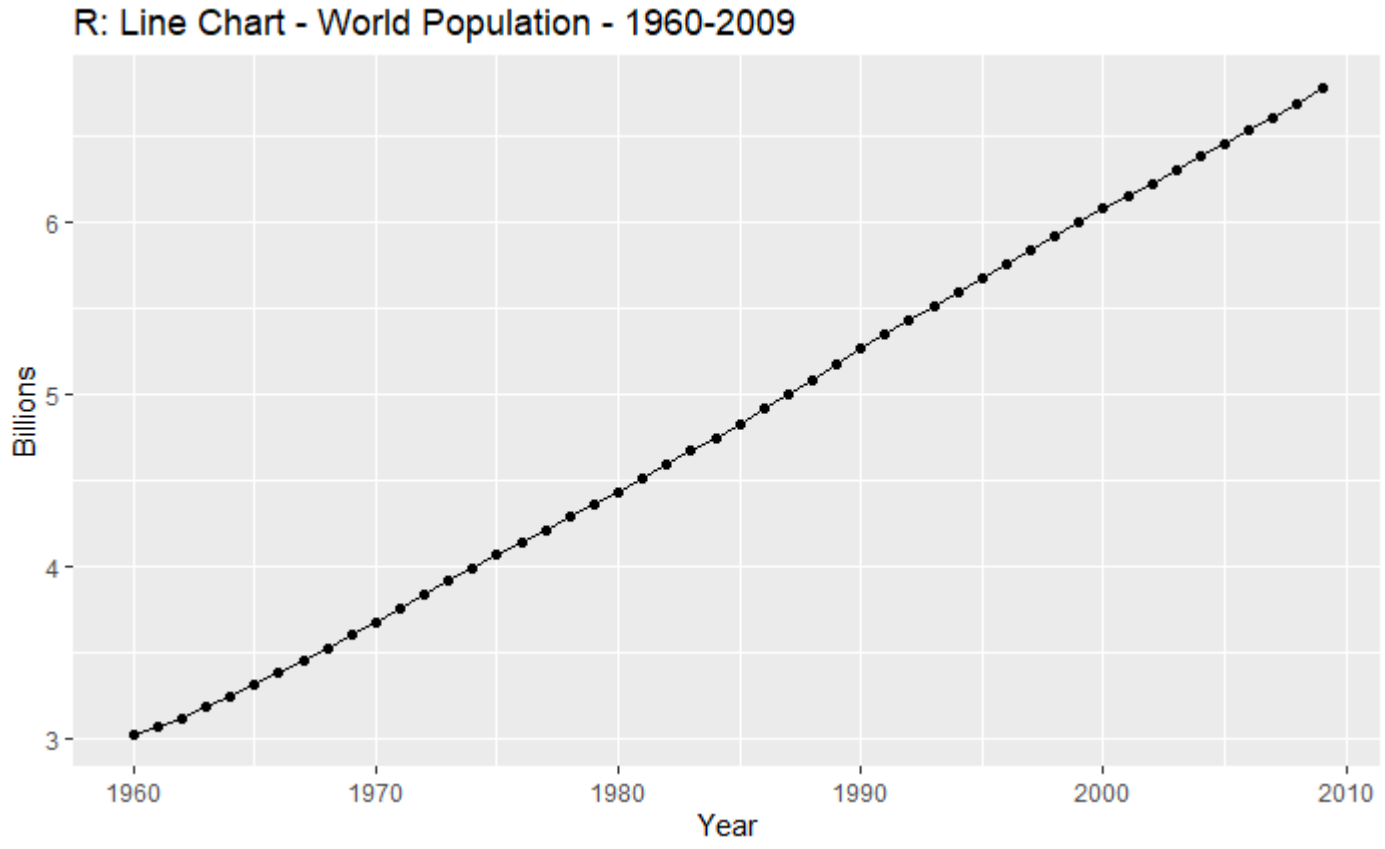
```
#import data
data = read_excel("C:\\Users\\longr\\Documents\\DSC 640\\Week 3 & 4\\2.2 Exercises\\world-popula
tion.xlsm")
```

Hide

```
data$PopinB <- round(data$Population/1000000000,2)
```

Hide

```
ggplot(data=data, aes(x=Year, y=PopinB, group=1)) +
  geom_line()+
  geom_point()+
  xlab("Year")+ylab("Billions")+
  ggtitle("R: Line Chart - World Population - 1960-2009")
```



Hide

```
ggplot(data=data, aes(x=Year, y=PopinB, group=1)) +
  geom_step()+
  geom_point()+
  xlab("Year")+ylab("Billions")+
  ggtitle("R: Step Chart - World Population - 1960-2009")
```