2 Graphical Description of Data

In chapter 1, you were introduced to the concepts of population, which again is a collection of all the measurements from the individuals of interest. Remember, in most cases you can't collect the entire population, so you have to take a sample. Thus, you collect data either through a sample or a census. Now you have a large number of data values. What can you do with them? No one likes to look at just a set of numbers. One thing is to organize the data into a table or graph. Ultimately though, you want to be able to use that graph to interpret the data, to describe the distribution of the data set, and to explore different characteristics of the data. The characteristics that will be discussed in this chapter and the next chapter are:

1. Center: middle of the data set, also known as the average.
2. Variation: how much the data varies.
3. Distribution: shape of the data (symmetric, uniform, or skewed).
4. Qualitative data: analysis of the data
5. Outliers: data values that are far from the majority of the data.
6. Time: changing characteristics of the data over time.

This chapter will focus mostly on using the graphs to understand aspects of the data, and not as much on how to create the graphs. There is technology that will create most of the graphs, though it is important for you to understand the basics of how to create them.

This textbook uses RStudio to perform all graphical and descriptive statistics, and all statistical inference. When using RStudio, every command is performed the same way. You start off with a goal(explanatory variable ~ response variable, data=data frame_name,...)

RStudio uses packages to make calculations easier. For this textbook, you will mostly need the package mosaic. There will be others that you will need on occasion, but you will be told that at the time. Most likely, mosaic is already installed in your RStudio. If you wish to install other packages you use the command

install.packages("name of package")

where you replace the name of package with the package you wish to install.

Once the package is installed, then you will need to tell RStudio you want to use it every time you start RStudio. The command to tell RStudio you want to use a package is

library("name of package")

You will need to turn on the package mosaic. The NHANES package contains a data frame that is useful. Both are accessed by running the command library("name of package").

Back to the basic command

goal(explanatory variable ~ response variable, data=data frame_name,...)

The goal depends on what you want to do. If you want to create a graph then you would need

gf_graph_type(explanatory_variable ~ response_variable, data=data_frame_name, ...)

As an example if you want to create a density plot of cholesterol levels on day 2 from a data frame called Cholesterol, then your command would be

gf_density(~day2, data=Cholesterol)

You will see more on what the different commands are that you would use. A word about the ... at the end of the command. That means there are other things you can do, but that is up to you if you want to actually do them. They do not need to be used if you don't want to. The following sections will show you how to create the different graphs that are usually completed in an introductory statistics course.

2.1 Qualitative Data

Remember, qualitative data are words describing a characteristic of the individual. There are several different graphs that are used for qualitative data. These graphs include bar graphs, Pareto charts, and pie charts. Bar graphs can be created using a statistical program like RStudio.

Bar graphs or charts consist of the frequencies on one axis and the categories on the other axis. Drawing the bar graph using r is performed using the following command.

gf_bar(~explanatory variable, data=Dataframe)

2.1.1 Example: Drawing a Bar Chart

Data was collected for two semesters in a statistics class. The data frame is in Table 2.1. The command

head(data frame)

shows the variables and the first few lines of the data set. The data sets are usually larger than what is shown. The head command allows one to see the structure of the data frame.

```
Class<-read.csv( "https://krkozak.github.io/MAT160/class_survey.csv")
knitr::kable(head(Class))
```

Table 2.1: Head of Statistics Class Survey

vehicle	gender	distance_campus	ice_cream	rent	major	height	winter
None	Female	1.5	Cookie Dough	724	Environmental and Sustainability Studies	61	Liked it
Mercury	Female	14.7	Sherbet	200	Administrative Justice	60	Don't like it
Ford	Female	2.4	Chocolate Brownie.	600	Bio Chem	68	Liked it
Toyota	Female	5.2	coffee	0		66	Loved it
Jeep	Male	2.0	Cookie Dough	600	Pre-health Careers	71	Loved it

vehicle	gender	distance_campus	ice_cream	rent	major	height	winter
Subaru	Male	5.0	none	500	Finance	72	No opinion

Every data frame has a code book that describes the data set, the source of the data set, and a listing and description of the variables in the data frame.

Code book for data frame class

Description Survey results from two semesters of statistics classes at Coconino Community College in the years 2018-2019.

Format

This data frame contains the following columns:

vehicle: Type of car a student drives

gender: Self declared gender of a student

distance_campus: how far a student lives from the Lone Tree Campus of Coconino Community College (miles)

ice_cream: favorite ice cream flavor

rent: How much a student pays in rent

major: Students declared major

height: height of the student (inches)

winter: Student's opinion of winter (Love it, Like it, Don't like, No opinion)

Source

Kozak K (2019). Survey results form surveys collected in statistics class at Coconino Community College.

References

Kozak, 2019

Create a bar graph of vehicle type. To do this in RStudio, use the command

gf_bar(~variable, data=Data_Frame, ...)

where gf_bar is the goal, vehicle is the name of the response variable (there is no explanatory variable), the data frame is Class, and a title was added to the graph.

2.1.1.1 Solution

```
gf_bar(~vehicle, data=Class, title="Bar Chart of Cars driven by students in statistics class", xl
```

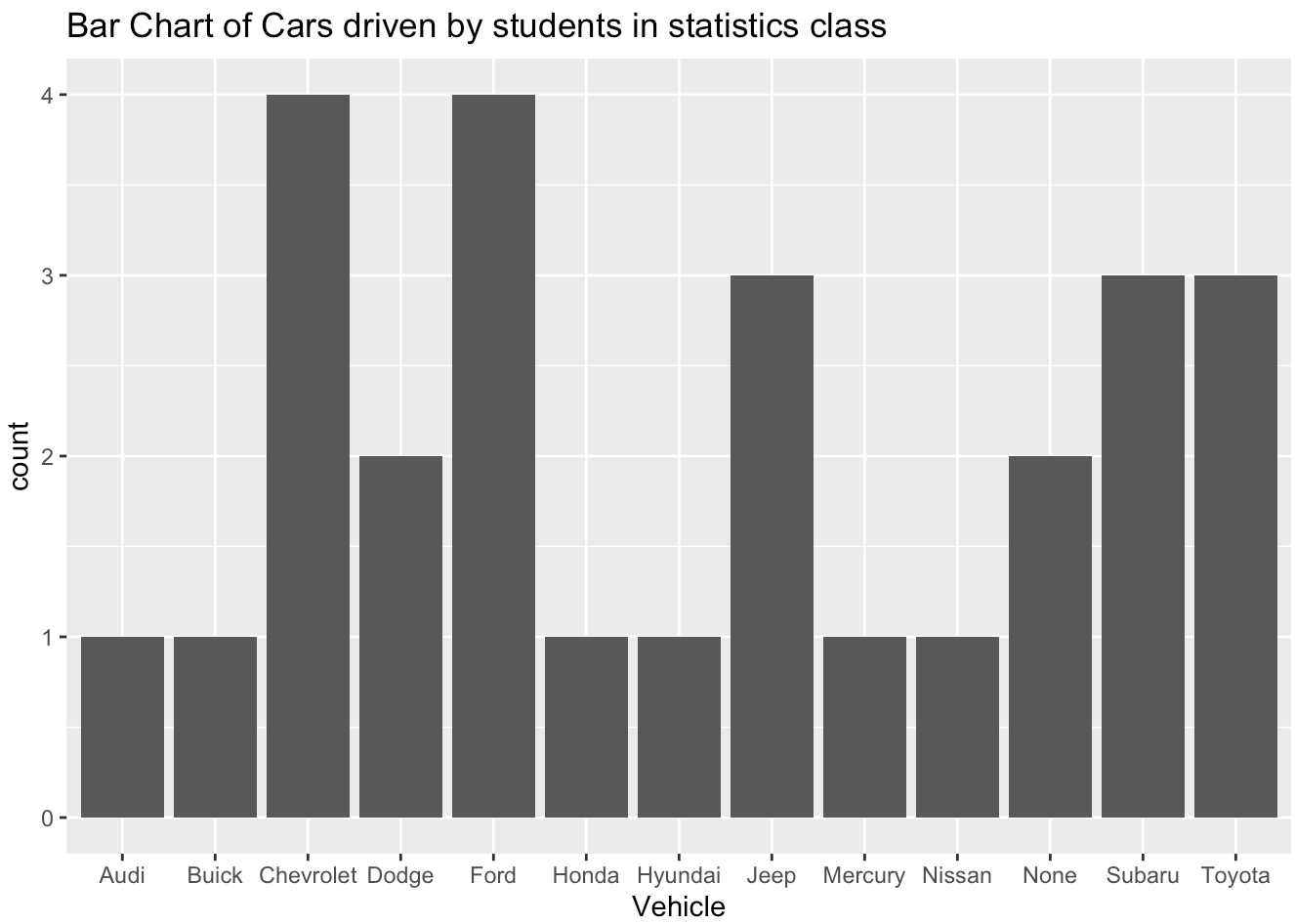


Figure 2.1: Cars driven by students in statistics class

Description of Figure 2.1 is a Bar graph with bars for Audi, Buick, Honda, Hyundai, Mercury, Nissan with height of 1, Dodge and None with height of 2, Jeep, Subaru, Toyota with heights of 3, and Chevrolet and Ford at height of 4.

Notice from Figure 2.1, you can see that Chevrolet and Ford are the more popular car, with Jeep, Subaru, and Toyota not far behind. Many types seems to be the lesser used, and tied for last place. However, more data would help to figure this out.

All graphs should have labels on each axis and a title for the graph.

The beauty of data frames with multiple variables is that you can answer many questions from the data. Suppose you want to see if gender makes a difference for the type of car a person drives. If you are a car manufacturer, if you knew that certain genders like certain cars, then you would advertise to the different genders. To create a bar graph that separates based on gender, perform the following command in RStudio.

```
gf_bar(~vehicle, fill=~gender, data=Class, title="Cars driving by students in statistics class",x
```

Figure 2.2: Bar graph of Cars driven by students in statistics class

Description of Figure 2.2 is a bar graph of number of vehicles separated by female and male. Audi and male has height of 1, Buick and female has a height of 1, Chevrolet and male and Chevrolet and female have heights of 2, Dodge and male and Dodge and female has heights of 1, Ford and female has a height of 4, Honda and female has a height of 1, Hyundai and male has a height of 1, Jeep and male has a height of 2 while Jeep and female has a height of 1, Mercury and female has a height of 1, Nissan and female has a height of 1, no car and female has a height of 2, Subaru and female has a height of 1, Subaru and male has a height of 2, Toyota and female has a height of 1, and Toyota and male has a height of 2.

Notice a Ford is driven by females more than any other car, while Chevrolet, Mercury, and Subaru cars are equally driven by males. Obviously a larger sample would be needed to make any conclusions from this data.

There are other types of graphs that can be created for quantitative variables. Another type is known as a dot plot. The command for this graph is as follows.

```
gf_dotplot(~vehicle, data=Class, title="Cars driven by students in statistics class", xlab="Vehic
```

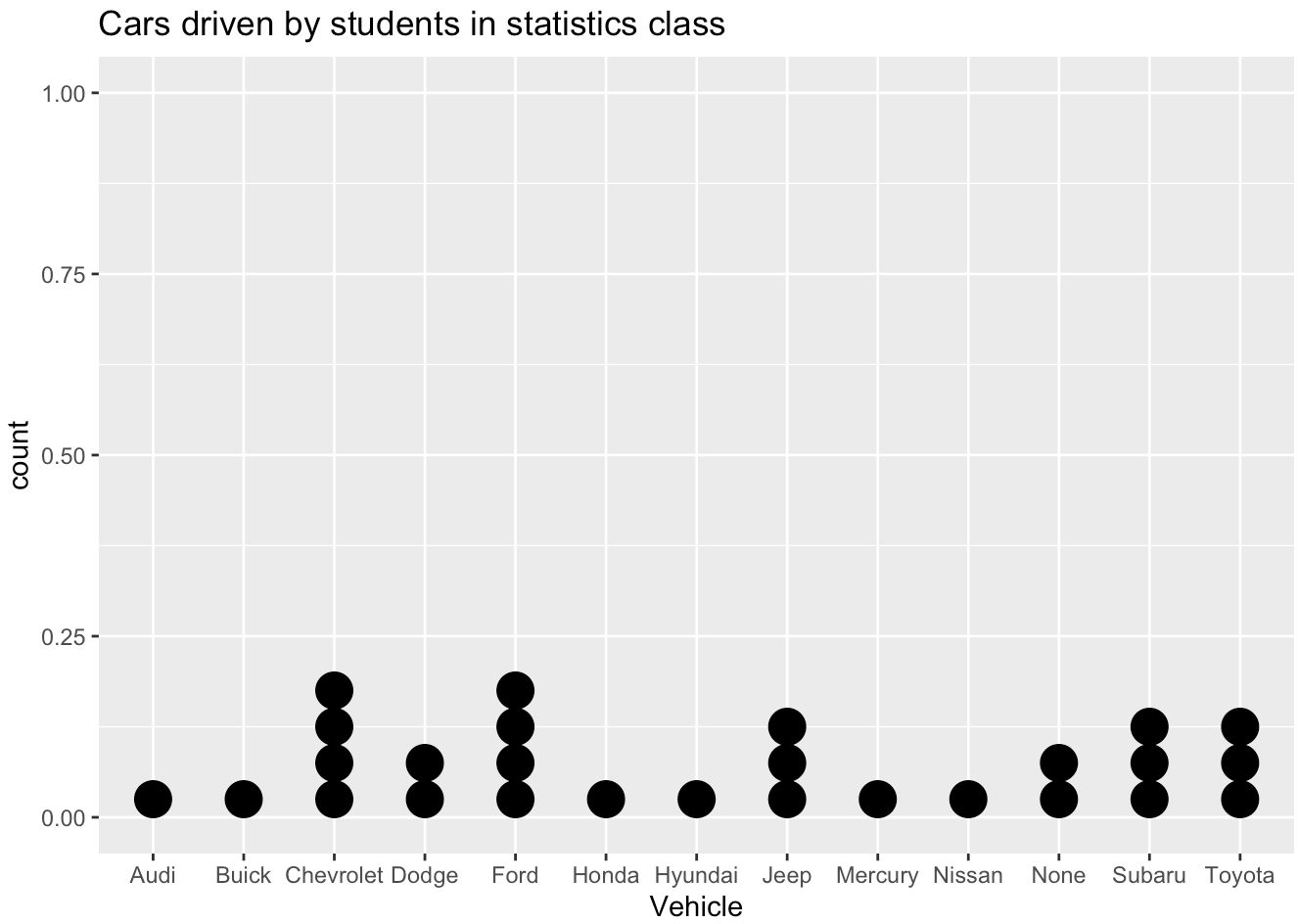


Figure 2.3: Cars driven by students in statistics class

Description of Figure 2.8 is a dot plot of number of vehicles with Audi, Buick, Honda, Hyundai, Mercury, Nissan with height of 1, Dodge and None with height of 2, Jeep, Subaru, Toyota with heights of 3, and Chevrolet and Ford at height of 4. Very similar to bar graph.

Notice a dot plot is like a bar chart. Both give you the same information. You can also divide a dot plot by gender.

Another type of graph that is also useful and similar to the dot plot is a point plot (scatter plot). In this plot you can graph the explanatory variable versus the response variable. The command for this in rStudio is as follows.

```
gf_point(vehicle~gender, data=Class, title="Cars driving by students in statistics class", xlab="(
```

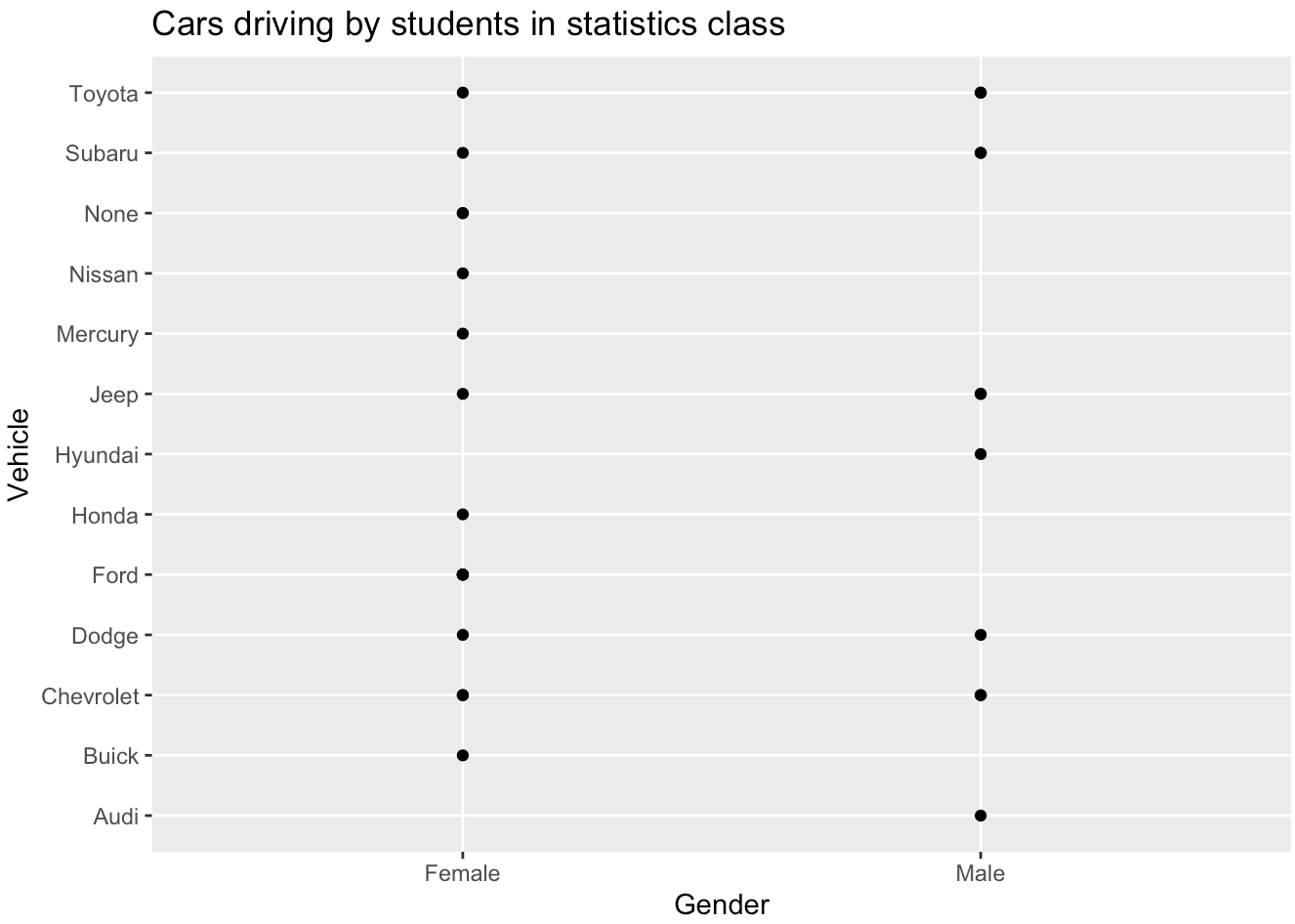


Figure 2.4: Cars driven by students in statistics class

Description of Figure 2.4 is a scatter plot of type of vehicles separated by female and male with females owning Toyota, Subaru, none, Nissan, Mercury, Jeep, Honda, Ford, Dodge, Chevrolet, and Buick, while males own Toyota, Subaru, Jeep, Hyundai, Dodge, Chevrolet, and Audi.

The problem with Figure 2.4 is that if there are multiple females who drive a Ford, only one dot is shown. So it is best to spread the dots out using a plot known as a jitter plot. In a jitter plot the dots are randomly moved off the center line. The command for a jitter plot is as follows:

```
gf_jitter(vehicle~gender, data=Class, title="Cars driving by students in statistics class", xlab=
```

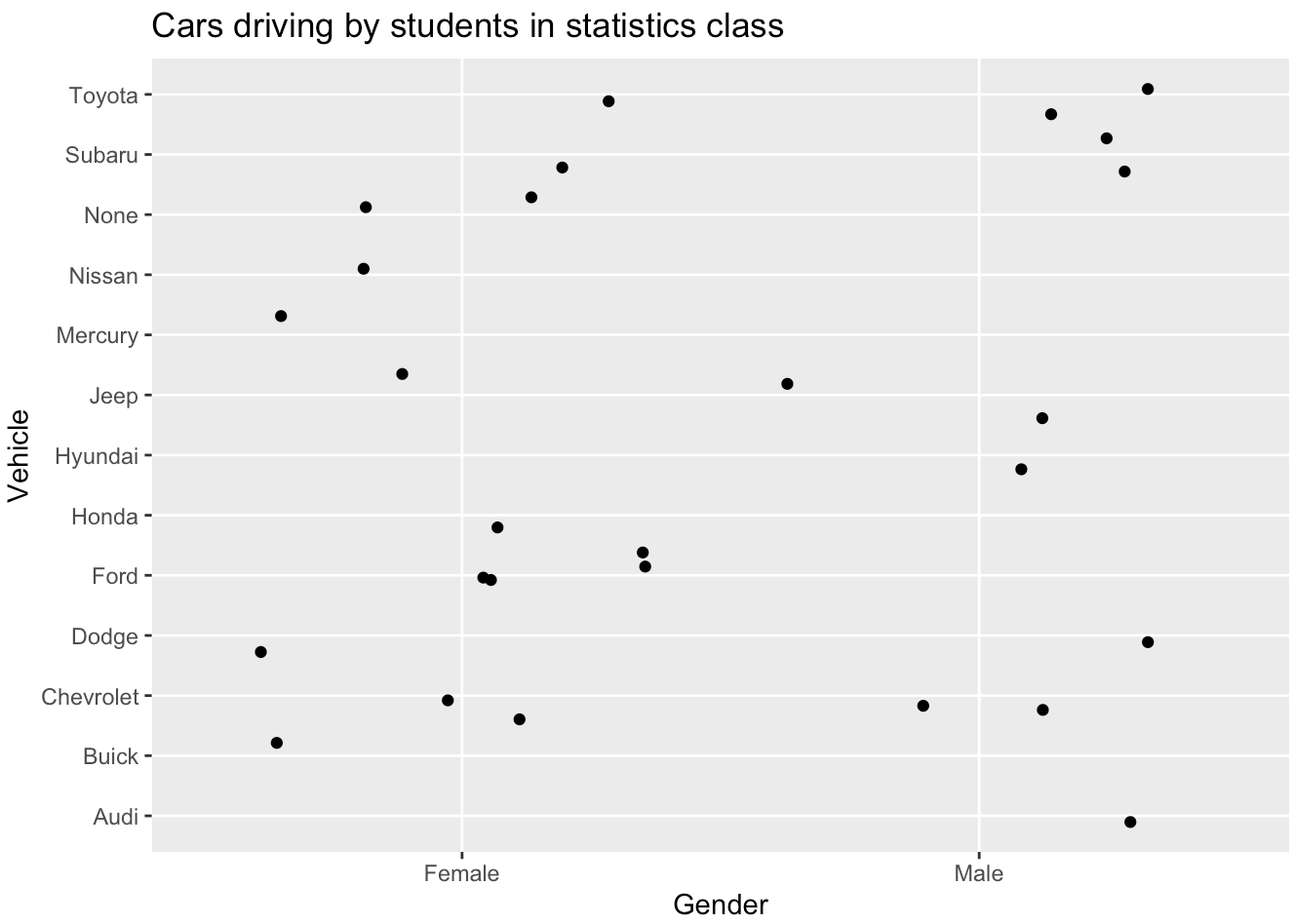


Figure 2.5: Cars driven by students in statistics class

Description of Figure 2.5 is a jitter plot of number of vehicles separated by female and male with females owning 1 Toyota, 1 Subaru, 2 with none, 1 Nissan, 1 Mercury, 1 Jeep, 1 Honda, 4 Fords, 1 Dodge, 2 Chevrolets, and 1 Buick, while males own 2 Toyotas, 2 Subarus, 2 Jeeps, 1 Hyundai, 1 Dodge, 1 Chevrolets, and 1 Audi.

Now you can observe that there are 4 females who drive a Ford. There is one female who drives a Honda. Other information about other cars and genders can be seen better than in the point plot and the bar graph. Jitter plots are useful to see how many data values are for each qualitative data values.

There are many other types of graphs that can be used on qualitative data. There are spreadsheet software packages that will create most of them, and it is better to look at them to see how to create then. It depends on your data as to which may be useful, but the bar, dot, and jitter plots are really the most useful.

2.1.2 Homework for Qualitative Data Section

1. Eyeglassomatic manufactures eyeglasses for different retailers. The number of lenses for different activities is in Table 2.2.

```
Eyeglasses<-read.csv( "https://krkozak.github.io/MAT160/eyglasses.csv")
knitr::kable(head(Eyeglasses))
```

Table 2.2: Head of Eyeglasses Data frame

activity
Grind
Grind
Grind
Grind
Grind
Grind

Code book for Data Frame Eyeglasses

Description Activities that an Eyeglass company performs when making eyeglasses, Grind means ground the lenses and put them in frames, multicoat means put tinting or coatings on lenses and then put them in frames, assemble means received frames and lenses from other sources and put them together, make frames means made the frames and put lenses in from other sources, receive finished means received glasses from other source unknown means do not know where the lenses came from.

Format

This data frame contains the following columns:

activity: The activity that is completed to make the eyeglasses by Eyeglassomatic

Source John Matic provided the data from a company he worked with. The company's name is fictitious, but the data is from an actual company.

References John Matic (2013)

Make a bar chart of this data. State any findings you can see from the graph.

2. Data was collected for two semesters in a statistics class drive. The data frame is in Table 2.1.

Code book for the Data Frame Class is found below Table 2.1.

Create a bar graph of the variable ice cream. State any findings you can see from the graphs.

3. The number of deaths in the US due to carbon monoxide (CO) poisoning from generators from the years 1999 to 2011 are in Table 2.3 (Hinatov, 2012). Create a bar chart of this data. State any findings you see from the graph.

```
Area<-read.csv( "https://krkozak.github.io/MAT160/area.csv")
knitr::kable(head(Area))
```

Table 2.3: Head of Area Data frame

deaths

deaths
Urban
Urban
Urban
Urban
Urban
Urban

4. Data was collected for two semesters in a statistics class drive. The data frame is in Table 2.1. Create a bar graph and dot plot of the variable major. Create a jitter plot of major and gender. State any findings you can see from the graphs.

   Code book for the Data Frame Class is found below Table 2.1.

5. Eyeglassomatic manufactures eyeglasses for different retailers. They test to see how many defective lenses they made during the time period of January 1 to March 31. The table Table 2.4 gives the defect and the number of defects. Create a bar chart of the data and then describe what this tells you about what causes the most defects.

```
Defects<- read.csv( "https://krkozak.github.io/MAT160/defects.csv")
knitr::kable(head(Defects))
```

Table 2.4: Head of Defects Data frame

type
small
small
pd
flaked
scratch
spot

Code book for Data Frame Defects

Description Types of defects that an Eyeglass company sees in the lenses they make into eyeglasses.

Format

This data frame contains the following columns:

type: The type of defect that is Seen when making eyeglasses by Eyeglassomatic

Source John Matic provided the data from a company he worked with. The company's name is fictitious, but the data is from an actual company.

References John Matic (2013)

6. American National Health and Nutrition Examination (NHANES) surveys is collected every year by the US National Center for Health Statistics (NCHS). The data frame is in Table 2.5. Create a bar chart of MartialStatus. Create a jitter plot of MaritalStatus versus Education. Describe any findings from the graphs.

```
knitr::kable(head(NHANES))
```

ID	SurveyYr	Gender	Age	AgeDecade	AgeMonths	Race1	Race3	Education	MaritalStatus	HHIncome	HHInco
51624	2009_10	male	34	30-39	409	White	NA	High School	Married	25000-34999	
51624	2009_10	male	34	30-39	409	White	NA	High School	Married	25000-34999	
51624	2009_10	male	34	30-39	409	White	NA	High School	Married	25000-34999	
51625	2009_10	male	4	0-9	49	Other	NA	NA	NA	20000-24999	
51630	2009_10	female	49	40-49	596	White	NA	Some College	LivePartner	35000-44999	
51638	2009_10	male	9	0-9	115	White	NA	NA	NA	75000-99999	

To view the code book for NHANES, type help("NHANES") in rStudio after you load the NHANES packages using library("NHANES")

2.2 Quantitative Data

There are several different graphs for quantitative data. With quantitative data, you can talk about how the data is distributed, called a distribution. The shape of the distribution can be described from the graphs.

Histogram: a graph of frequencies (counts) on the vertical axis and classes on the horizontal axis. The height of the rectangles is the frequency and the width is the class width. The width depends on how many classes (bins) are in the histogram. The shape of a histogram is dependent on the number of bins. In RStudio the command to create a histogram is

gf_histogram(~response variable, data=Data_Frame, title="title of the graph")

The last part of the command puts a title on the graph. You type in what ever you want for the title in the quotes.

Density Plot: Similar to a histogram, except smoothing is created to smooth out the graph. The shape is not dependent on the number of bins so the distribution is easier to determine from the density plot. In RStudio the command to create a density plot is

gf_density(~response variable, data=Data_Frame, title="title of the graph", xlab="Label", ylab="Label")

The last part of the command puts a title on the graph and labels on the axes. You type in what every you want for the title and labels in the quotes.

The last part of the command puts a title on the graph and labels on the axes. You type in what every you want for the title and labels in the quotes.

2.2.1 Example: Drawing a Histogram and Density plot

Data was collected for two semesters in a statistics class drive. The data frame is in Table 2.1 and the code book is below the data frame

Draw a histogram, density plot, and a dot plot for the variable the distance a student lives from the Lone Tree Campus of Coconino Community College. Describe the story the graphs tell.

2.2.1.1 Solution

```
gf_histogram(~distance_campus, data=Class, title="Distance in miles from the Lone Tree Campus", x
```

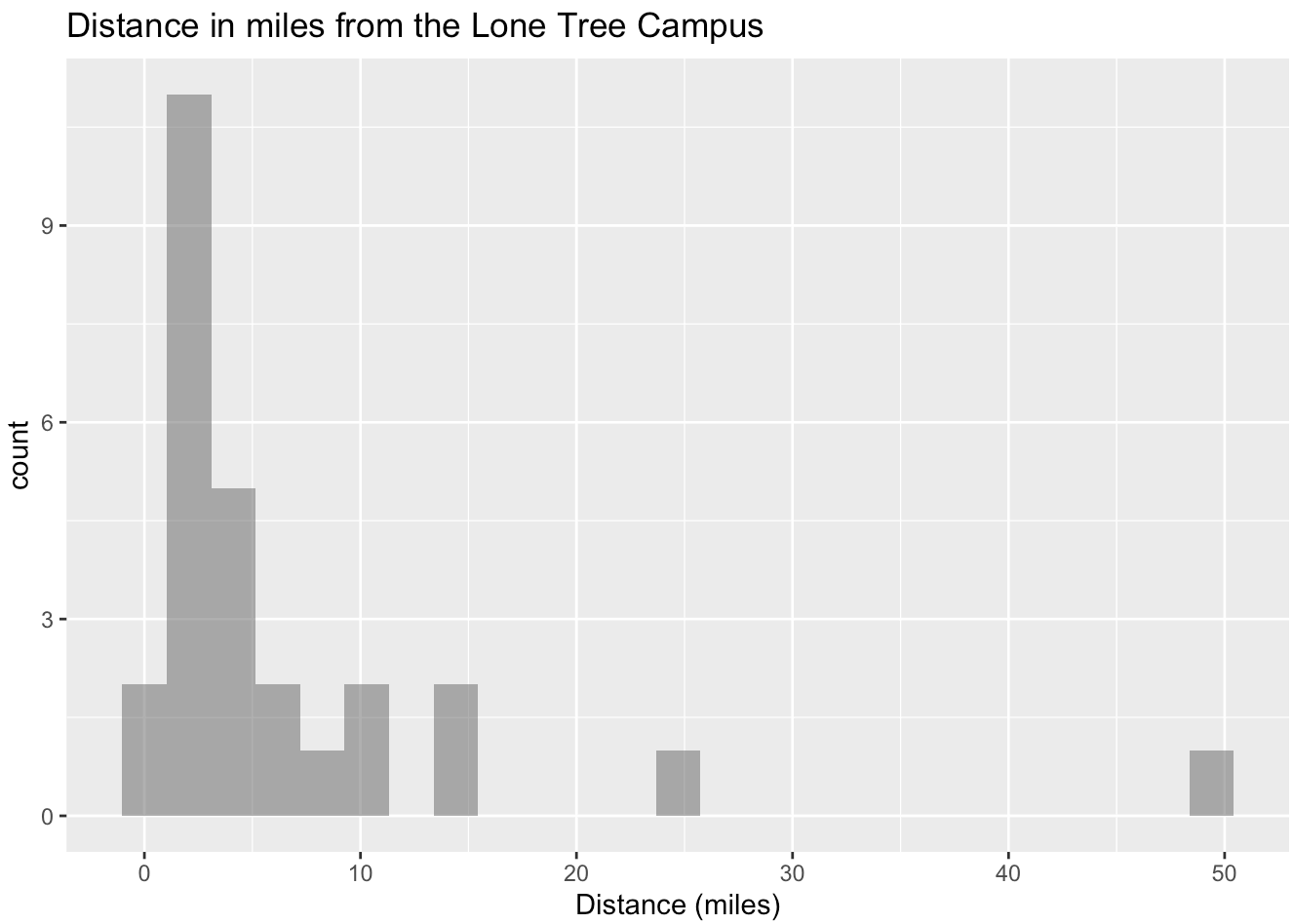


Figure 2.6: Distance in miles from the Lone Tree Campus

Description of the graph is histogram with high part on left and low part on right with several gaps. The graph contains bars.

```
gf_density(~distance_campus, data=Class, title="Distance in miles from the Lone Tree Campus", xlal
```

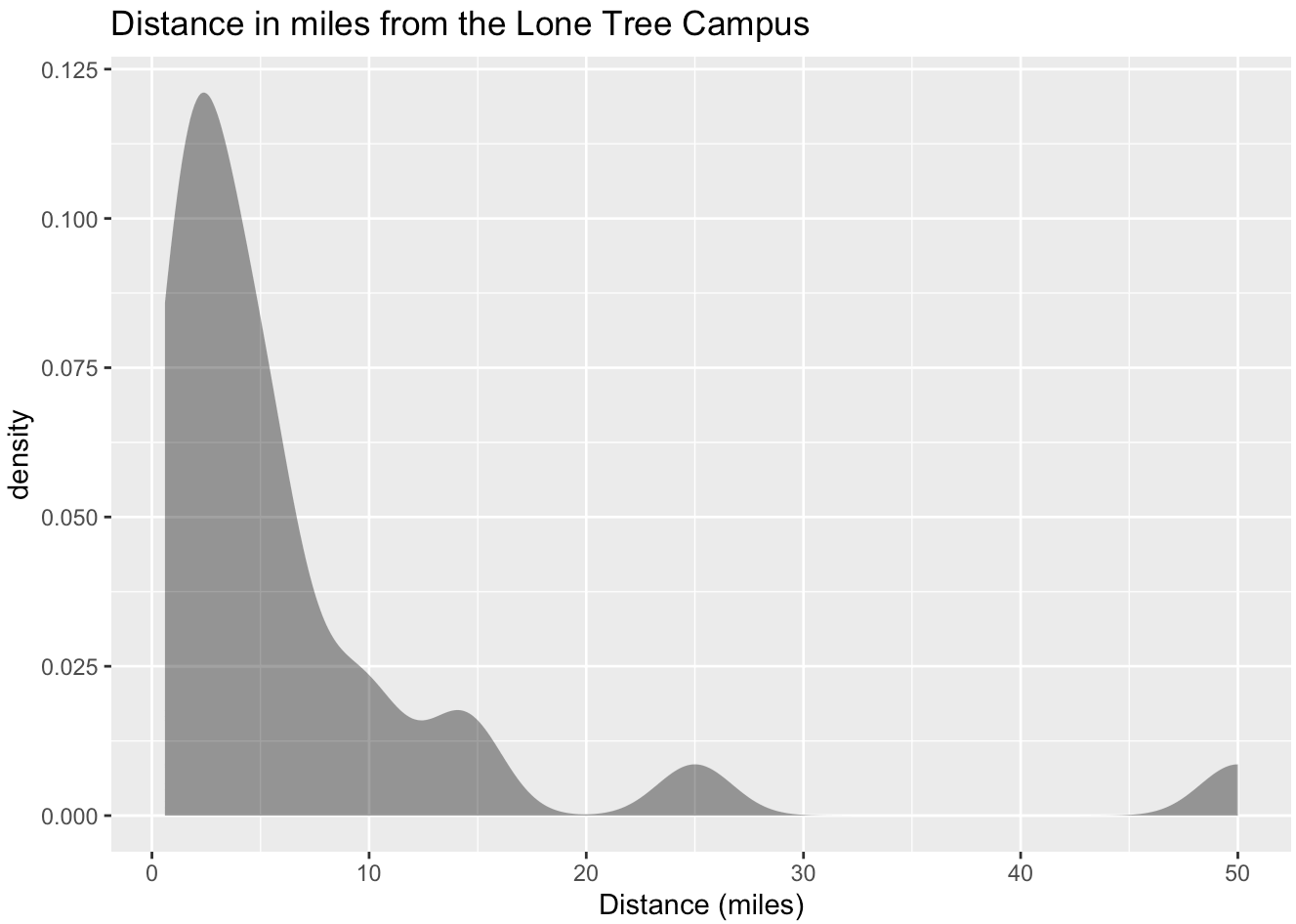


Figure 2.7: Distance in miles from the Lone Tree Campus

Description of the graph is density graph with high part on left and low part on right with several gaps. The graph is smooth.

```
gf_dotplot(~distance_campus, data=Class, title="Distance in miles from the Lone Tree Campus", xlal
```

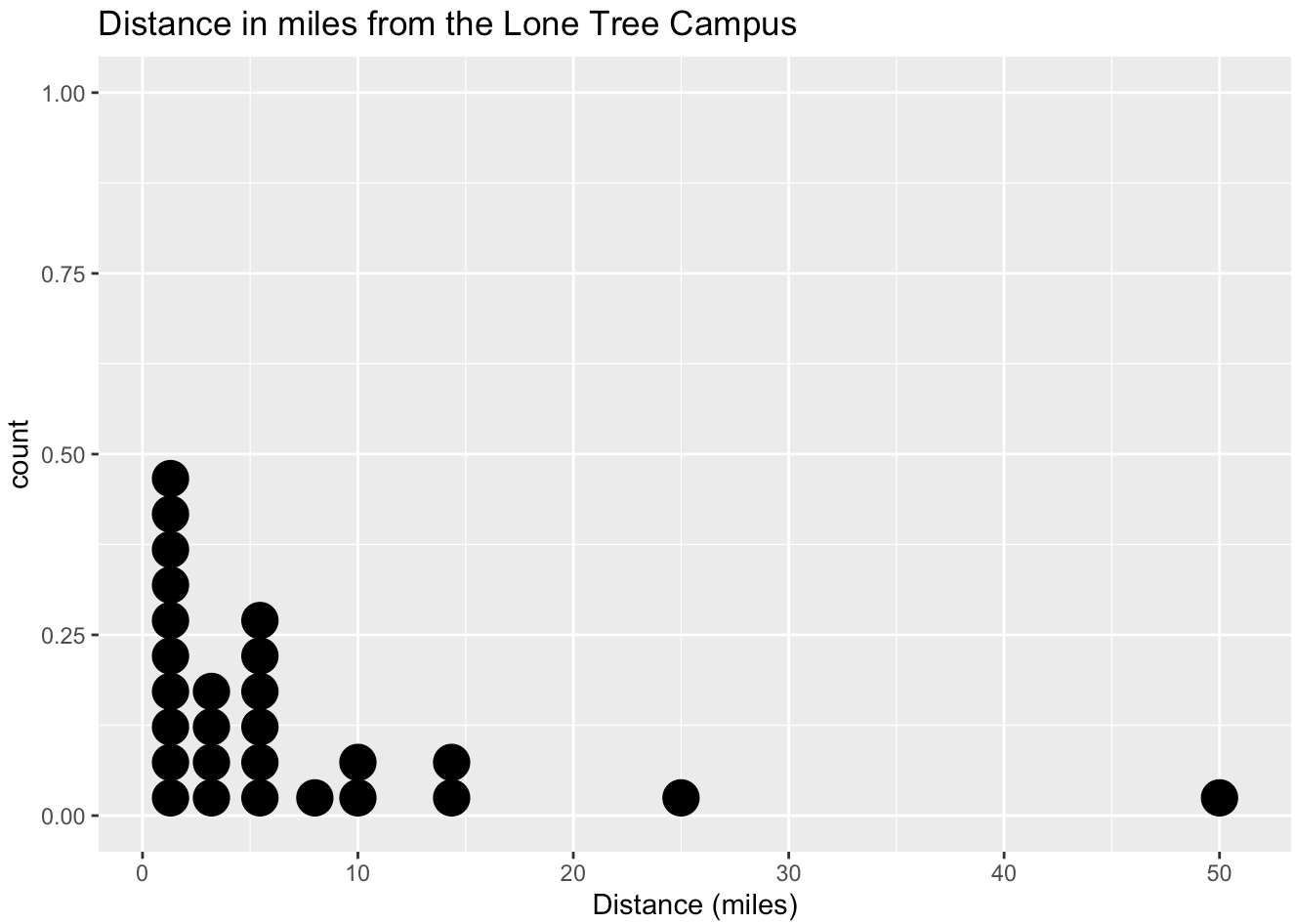


Figure 2.8: Distance in miles from the Lone Tree Campus

Description of the graph of dot plot with high part on left and low part on right with several gaps. The graph is with dots that represent each data value.

Notice the histogram, density plot, and dot plot are all very similar, but the density plot is smoother. They all tell you similar ideas of the shape of the distribution. Reviewing the graphs you can see that most of the students live within 10 miles of the Lone Tree Campus, in fact most live within 5 miles from the campus. However, there is a student who lives around 50 miles from the Lone Tree Campus. This is a great deal farther from the rest of the data. This value could be considered an outlier. An outlier is a data value that is far from the rest of the values. It may be an unusual value or a mistake. It is a data value that should be investigated. In this case, the student lived really far from campus, thus the value is not a mistake, and is just very unusual. The density plot is probably the best plot for most data frames.

There are other aspects that can be discussed, but first some other concepts need to be introduced.

2.2.2 Shapes of the distribution:

When you look at a distribution, look at the basic shape. There are some basic shapes that are seen in histograms. Realize though that some distributions have no shape. The common shapes are symmetric, skewed, and uniform. Another interest is how many peaks a graph may have. This is known as modal.

Symmetric means that you can fold the graph in half down the middle and the two sides will line up. You can think of the two sides as being mirror images of each other. Skewed means one “tail” of the graph is

longer than the other. The graph is skewed in the direction of the longer tail (backwards from what you would expect). A uniform graph has all the bars the same height.

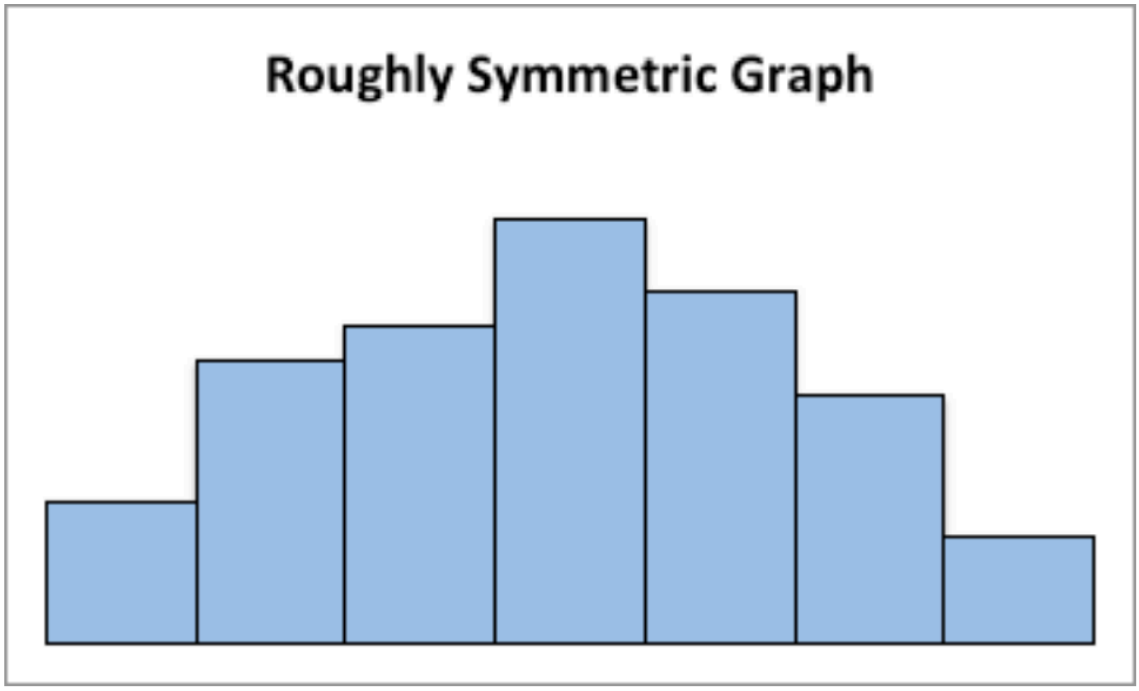
Modal refers to the number of peaks. Unimodal has one peak and bimodal has two peaks. Usually if a graph has more than two peaks, the modal information is not longer of interest.

Other important features to consider are gaps between bars, a repetitive pattern, how spread out is the data, and where the center of the graph is.

2.2.3 Examples of graphs:

This graph is roughly symmetric and unimodal:

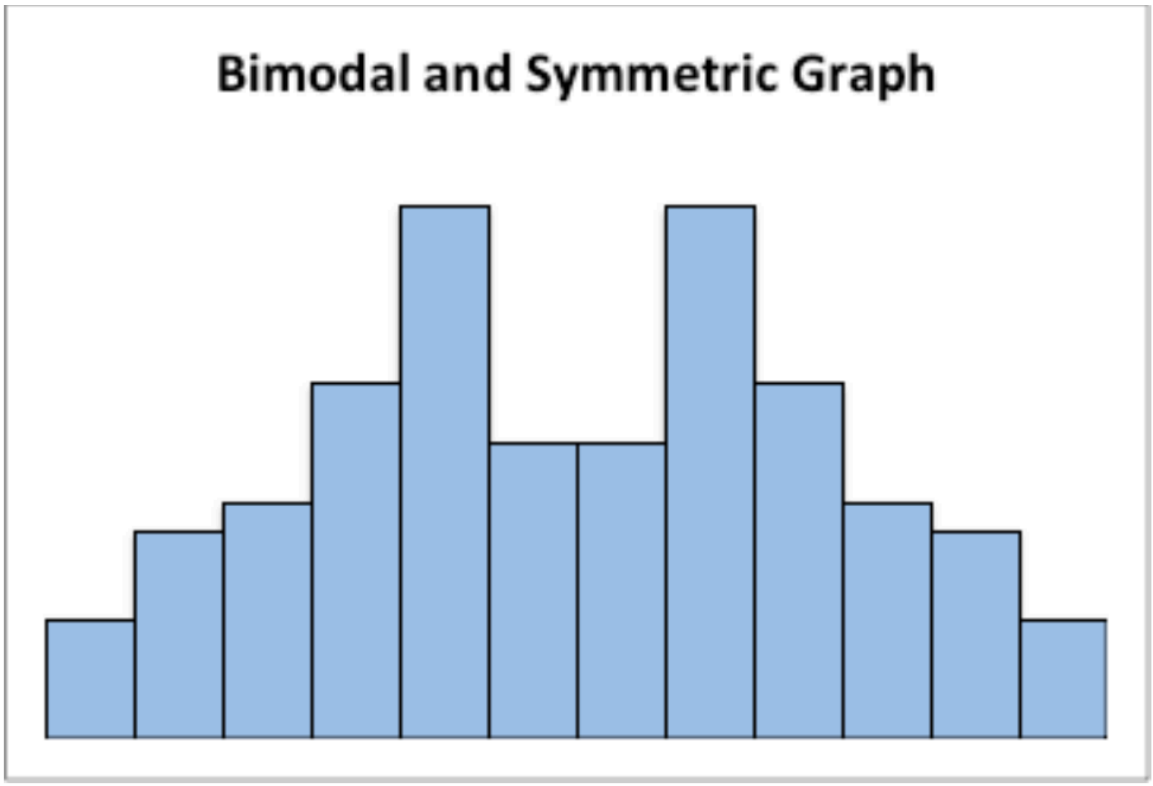
Graph: Symmetric Distribution



symmetric Graph

This graph is symmetric and bimodal:

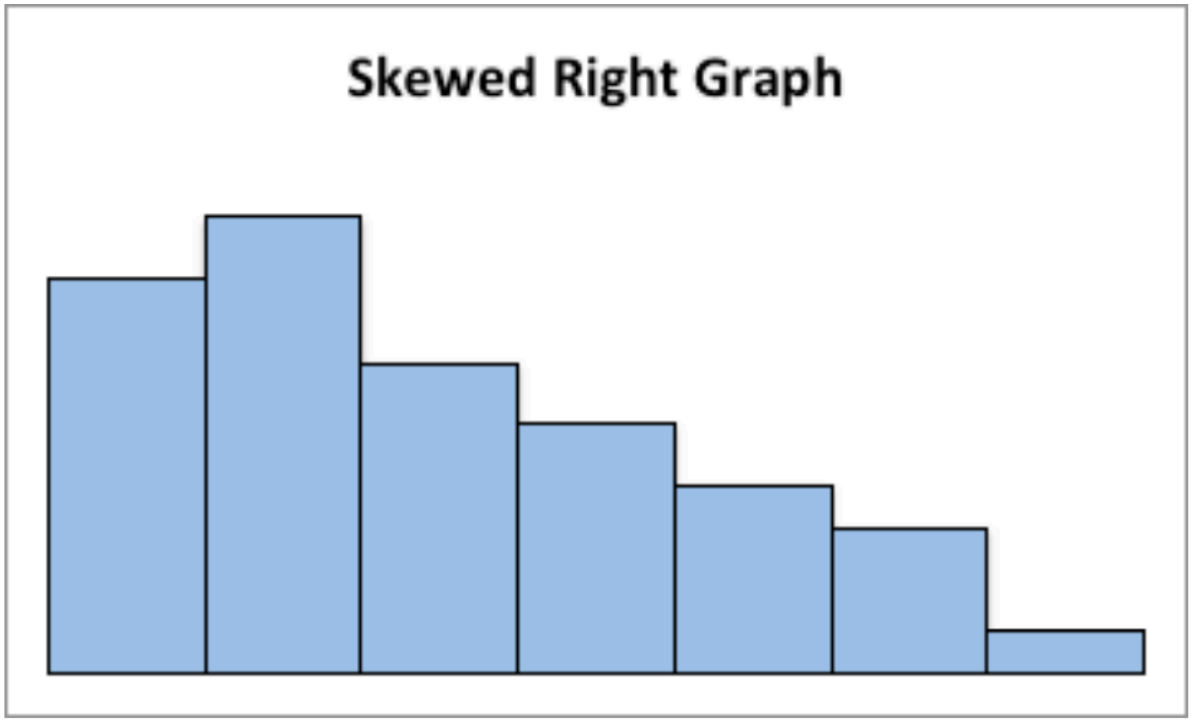
Graph: Symmetric and Bimodal Distribution



Bimodal and symmetric graph

This graph is skewed to the right:

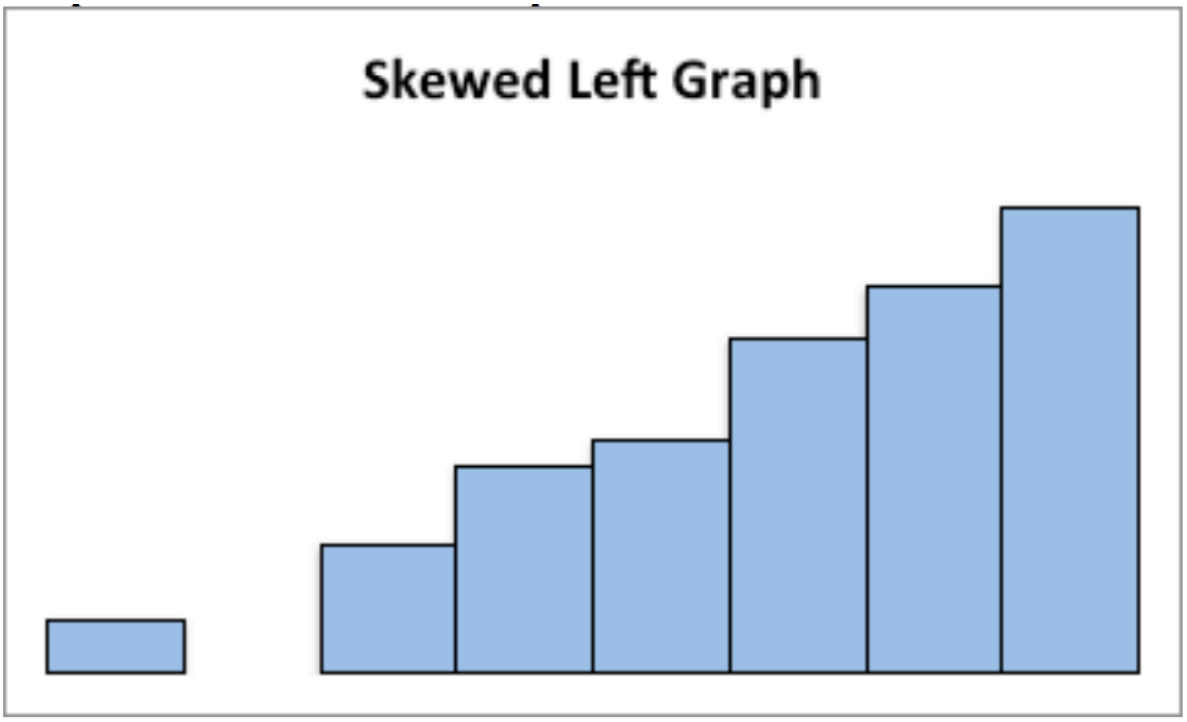
Graph: Skewed Right Distribution



Skewed right graph

This graph is skewed to the left and has a gap:

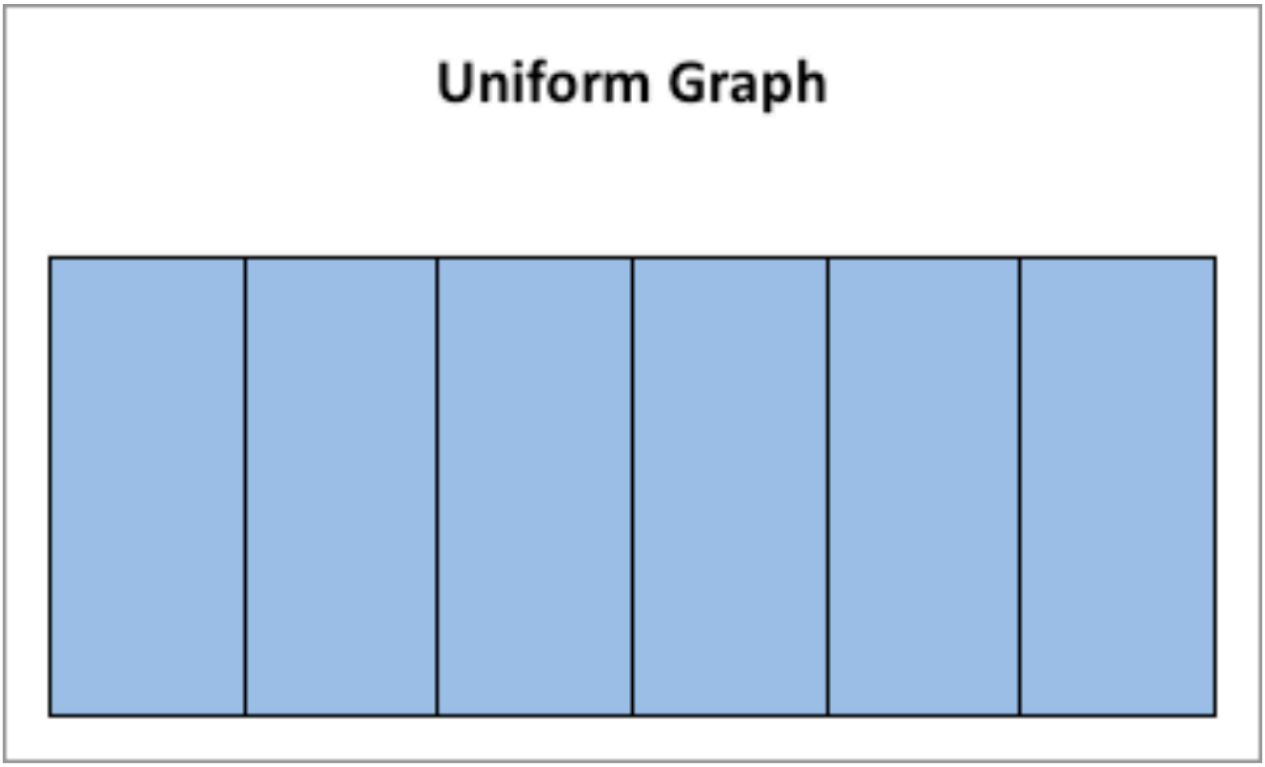
Graph: Skewed Left Distribution



Skewed Left graph

This graph is uniform since all the bars are the same height:

Graph: Uniform Distribution



Uniform graph

2.2.4 Example: Drawing a Histogram and Density plot

Data was collected from the Chronicle of Higher Education for tuition from public four year colleges, private four year colleges, and for profit four year colleges. The data frame is in Table 2.6. Draw a density plot of instate tuition levels for all four year institutions, and then separate the density plot for instate tuition based on type of institution. Describe any findings from the graph.

```
Tuition<-read.csv( "https://krkozak.github.io/MAT160/Tuition_4_year.csv")
knitr::kable(head(Tuition))
```

Table 2.6: Head of Tuition Data Frame

INSTITUTION	TYPE	STATE	ROOM_BOARD	INSTATE_TUITION	INSTATE_TOTAL	OUTOFSTATE_TUITION
University of Alaska AnchoragePublic 4-year	Public_4 year	AK	12200	7688	19888	23858
University of Alaska FairbanksPublic 4-year	Public_4 year	AK	8930	8087	17017	24257
University of Alaska SoutheastPublic 4-year	Public_4 year	AK	9200	7092	16292	19404
Alaska Bible CollegePrivate 4-year	Private_4_year	AK	5700	9300	15000	9300
Alaska Pacific UniversityPrivate 4-year	Private_4_year	AK	7300	20830	28130	20830
Alabama Agricultural and Mechanical UniversityPublic 4-year	Public_4 year	AL	8379	9698	18077	17918

Code book for Data Frame Tuition

Description Cost of four year institutions.

Format

This data frame contains the following columns:

INSTITUTION: Name of four year institution

TYPE: Type of four year institution, Public_4_year, Private_4_year, For_profit_4_year.

STATE: What state the institution resides

ROOM_BOARD: The cost of room and board at the institution (\$)

INSTATE_TUTION: The cost of instate tuition (\$)

INSTATE_TOTAL: The cost of room and board and instate tuition (\$ per year)

OUTOFSTATE_TUTION: The cost of out of state tuition (\$ per year)

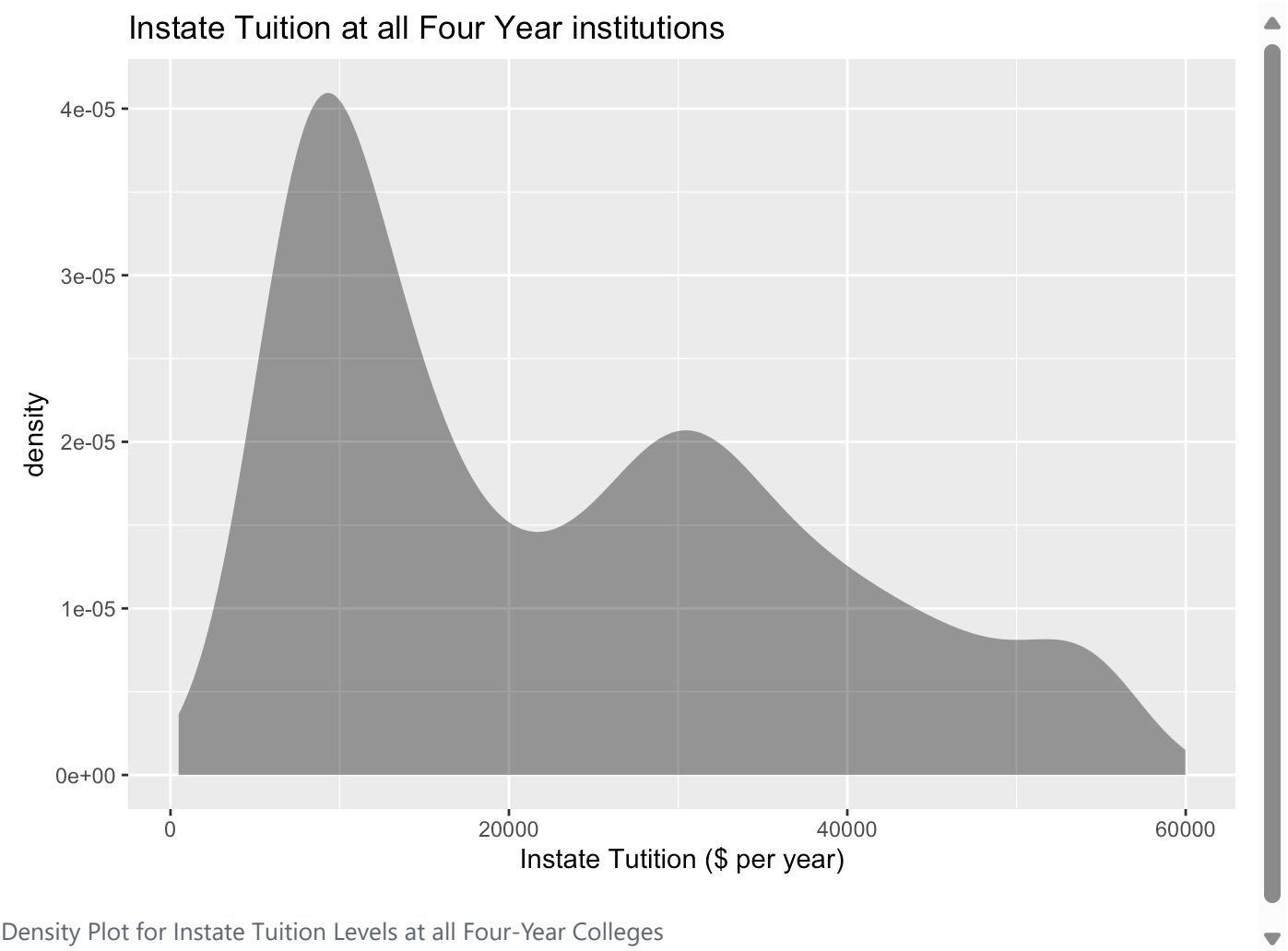
OUTOFSTATE_TOTAL: The cost of room and board and out of state tuition (\$ per year)

Source Tuition and Fees, 1998-99 Through 2018-19. (2018, December 31). Retrieved from https://www.chronicle.com/interactives/tuition-and-fees

References Chronicle of Higher Education *, December 31, 2018.

2.2.4.1 **Solution**

```
gf_density(~INSTATE_TUITION, data=Tuition, title="Instate Tuition at all Four Year institutions",
```



Density Plot for Instate Tuition Levels at all Four-Year Colleges

Description of the graph is a density with high part on left, then a dip and up to peak in the middle that is lower than the left peak and then the lowest peak on the right .

(ref:tuition-instate-type-cap) Density Plot for Instate Tuition Levels at all Four-Year Colleges

```
gf_density(~INSTATE_TUITION|TYPE, data=Tuition, title="Instate Tuition at all Four Year institions
```

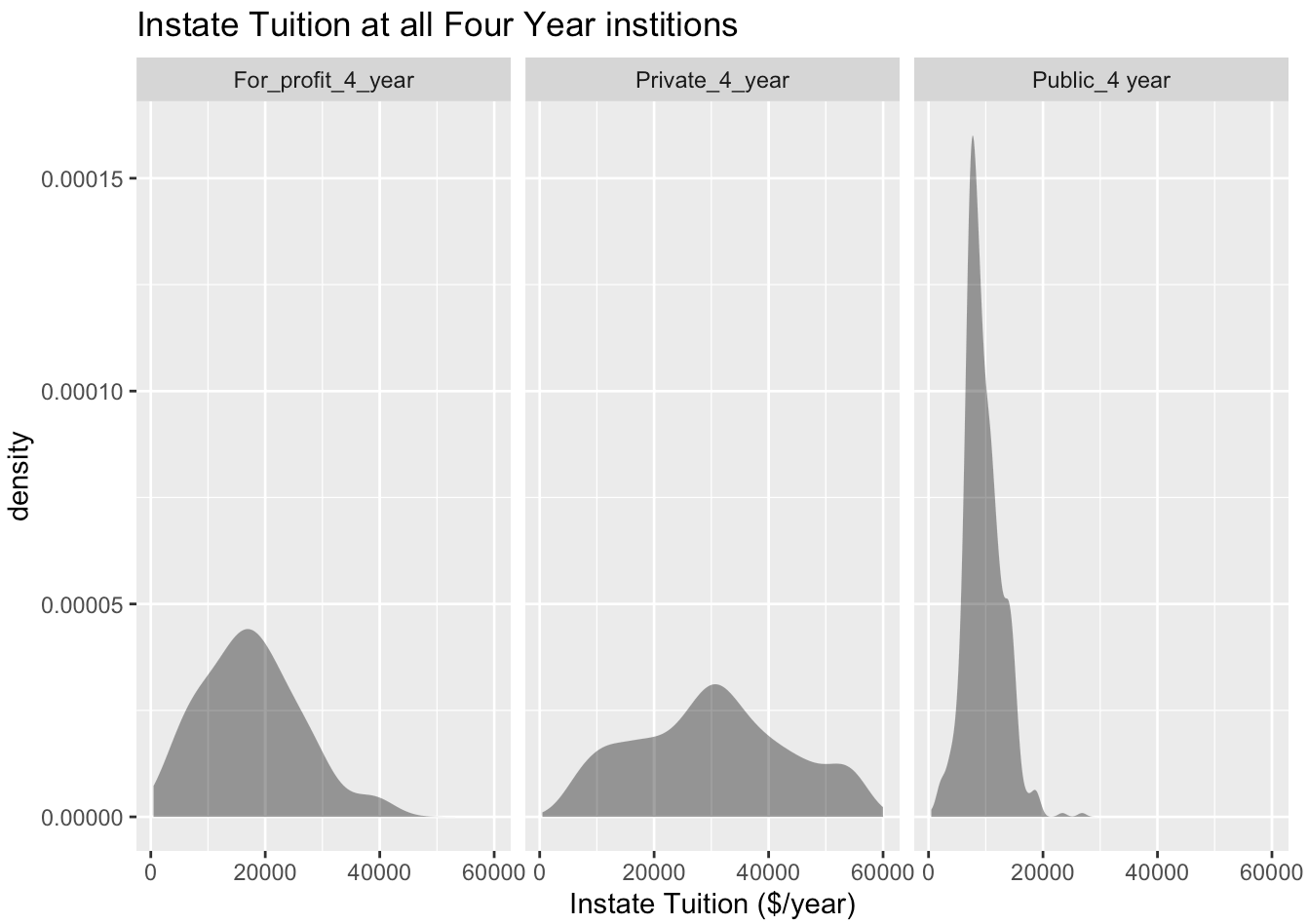


Figure 2.9: Instate Tuition at all Four Year institions

Description of Figure 2.9 is a density plots separated by for profit 4 year with peak on left, private 4 year with peak in the middle, and public 4 year colleges with peak on the left. Public 4 year has the highest peak, with for profit 4 year is lower, and then private 4 year with the lowest peak.

The distribution is skewed right, with no gaps. Most institutions in state is less than \$ 20,000 per year though some go as high as \$ 60,000 per year. When separated by public versus private and for profit, most public are much less than \$ 20,000 per year while private four year cost around \$ 30,000 per year, and for profit are around \$ 20,000 per year.

There are other types of graphs for quantitative data. They will be explored in the next section.

2.2.5 Homework for Quantitative Data Section

1. The weekly median incomes of males and females for specific occupations, are given in Table 2.7 (CPS News Releases. (n.d.). Retrieved July 8, 2019, from https://www.bls.gov/cps/). Create a density plot for males and females. Discuss any findings from the graph. Note: to put two graphs on the same axis, type the piping symbol |> (base r) or %>% (magrittr package) (Note: |> and %>% are piping symbols that can be thought of as "and then") at the end of the first command and then type the command for the second graph on the next line. Also, use fill="pick a color" in the command to plot the graphs with different colors so the two graphs can be easier to distinguish.

```
Wages<- read.csv( "https://krkozak.github.io/MAT160/wages.csv")
knitr::kable(head(Wages))
```

Table 2.7: Head of Wages Data frame

Occupation	Numworkers	median_wage	male_worker	male_wage	female_worker	female_wage
Management, professional, and related occupations	48808	1246	23685	1468	25123	1078
Management, business, and financial operations occupations	19863	1355	10668	1537	9195	1168
Management occupations	13477	1429	7754	1585	5724	1236
Chief executives	1098	2291	790	2488	307	1736
General and operations managers	939	1338	656	1427	283	1139
Legislators	14	NA	10	NA	4	NA

Code book for Data Frame Wages

Description Median weekly earnings of full-time wage and salary workers by detailed occupation and sex. The Current Population Survey (CPS) is a monthly survey of households conducted by the Bureau of Census for the Bureau of Labor Statistics. It provides a comprehensive body of data on the labor force, employment, unemployment, persons not in the labor force, hours of work, earnings, and other demographic and labor force characteristics.

Format

This data frame contains the following columns:

Occupation: Occupations of workers.

Numworkers: The number of workers in each occupation (in thousands of workers)

median_wage: Median weekly wage (\$)

male_worker: number of male workers (in thousands of workers)

male_wage: Median weekly wage of male workers (\$)

female_worker: number of female workers (in thousands of workers)

female_wage: Median weekly wage of female workers (\$)

Source CPS News Releases. (n.d.). Retrieved July 8, 2019, from https://www.bls.gov/cps/

References Current Population Survey (CPS) retrieved July 8, 2019.

2. The density of people per square kilometer for certain countries is in Table 2.8 (World Bank, 2019). Create density plot of density in 2018 for just Sub-Saharan Africa. Describe what story the graph tells.

```
Density<- read.csv( "https://krkozak.github.io/MAT160/density.csv")
knitr::kable(head(Density))
```

Country_Name	Country_Code	Region	IncomeGroup	y1961	y1962	y1963	y1964	y196
Aruba	ABW	Latin America & Caribbean	High income	307.988889	312.361111	314.972222	316.844444	318.66666
Afghanistan	AFG	South Asia	Low income	14.044987	14.323808	14.617537	14.926295	15.25031
Angola	AGO	Sub-Saharan Africa	Lower middle income	4.436891	4.498708	4.555593	4.600180	4.62867
Albania	ALB	Europe & Central Asia	Upper middle income	60.576642	62.456898	64.329234	66.209307	68.05806
Andorra	AND	Europe & Central Asia	High income	30.585106	32.702128	34.919149	37.168085	39.46595
Arab World	ARB			8.430860	8.663154	8.903441	9.152526	9.41096

Code book for Data Frame Density

Description Population density of all countries in the world

Format

This data frame contains the following columns:

Country_Name: The name of countries or regions around the world

Country_Code: The 3 letter code for a country or region

Region: World Banks classification of where the country is in the world

Incomegroup: World Banks classification of what income level the country is considered to be

y1961-y2018: population density for the years 1961 through 2018, people per sq. km of land area, population density is midyear population divided by land area in square kilometers. Population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship–except for refugees not permanently settled in the country of asylum, who are generally considered part of the population of their country of origin. Land area is a country's total area, excluding area under inland water bodies, national claims to continental shelf, and exclusive economic zones. In most cases the definition of inland water bodies includes major rivers and lakes.

Source Population density (people per sq. km of land area). (n.d.). Retrieved July 9, 2019, from https://data.worldbank.org/indicator/EN.POP.DNST

References Food and Agriculture Organization and World Bank population estimates.

Since the Density data frame is for all countries, a new data frame must be created with just Sub-Saharan Africa Table 2.9. This is created by using the following command

```
Africa <- Density |>
  filter(Region == "Sub-Saharan Africa")
knitr::kable(head(Africa))
```

Country_Name	Country_Code	Region	IncomeGroup	y1961	y1962	y1963	y1964	y
Angola	AGO	Sub-Saharan Africa	Lower middle income	4.4368910	4.4987078	4.5555932	4.6001797	4.628
Burundi	BDI	Sub-Saharan Africa	Low income	111.0762461	113.2134346	115.4371885	117.8461838	120.497
Benin	BEN	Sub-Saharan Africa	Low income	21.8682778	22.1966655	22.5510731	22.9333540	23.344
Burkina Faso	BFA	Sub-Saharan Africa	Low income	17.8895468	18.1298465	18.3765387	18.6362939	18.913
Botswana	BWA	Sub-Saharan Africa	Upper middle income	0.9046371	0.9242108	0.9452208	0.9667267	0.988
Central African Republic	CAF	Sub-Saharan Africa	Low income	2.4496228	2.4911073	2.5351857	2.5821310	2.632

3. The Affordable Care Act created a market place for individuals to purchase health care plans. In 2014, the premiums for a 27 year old for the different levels health insurance are given in Table 2.10 (\"Health insurance marketplace,\" 2013). Create a density plot of bronze_lowest, then silver_lowest, and gold_lowest all on the same aces. Use |> or %>% at the end of each command. Describe the story the graphs tells.

```
Insurance<- read.csv( "https://krkozak.github.io/MAT160/insurance.csv")
knitr::kable(head(Insurance))
```

Table 2.10: Head of Insurance D

state	average_QHP	bronze_lowest	silver_lowest	gold_lowest	catastrophic	second_silver_pretax	second_silver_po
AK	34	254	312	401	236	312	
AL	7	162	200	248	138	209	
AR	28	181	231	263	135	241	
AZ	106	141	164	187	107	166	
DE	19	203	234	282	137	237	
FL	102	169	200	229	132	218	

Code book for Data Frame Insurance

Description The Affordable Care Act created a market place for individuals to purchase health care plans.The data is from 2014.

Format

This data frame contains the following columns:

state: state of insured.

average_QHP: The number of qualified health plans

bronze_lowest: premium for the lowest bronze level of insurance for a single person (\$)

silver_lowest: premium for the lowest silver level of insurance for a single person (\$)

gold_lowest: premium for the lowest gold level of insurance for a single person (\$)

catastrophic: premium for the catastrophic level of insurance for a single person (\$)

second_silver_pretax: premium for the second silver level of insurance for a single person pretax (\$)

second_silver_posttax: premium for the second silver level of insurance for a single person posttax (\$)

second_bronze_posttax: premium for the lowest bronze level of insurance for a single person posttax (\$)

silver_family_pretax: premium for the silver level of insurance for a family pretax (\$)

silver_family_posttax: premium for the silver level of insurance for a family posttax (\$)

bronze_family_posttax: premium for the bronze level of insurance for a family posttax (\$)

Source Health Insurance Market Place Retrieved from website: http://aspe.hhs.gov/health/reports/2013/marketplacepremiums/ib_premiumslandscape.pdf premiums for 2014.

References Department of Health and Human Services, ASPE. (2013). Health insurance marketplace

4. Students in a statistics class took their first test. In Table 2.11 are the scores they earned. Create a density plot for grades. Describe the shape of the distribution.

```
Firsttest_1<- read.csv( "https://krkozak.github.io/MAT160/firsttest_1.csv")
knitr::kable(head(Firsttest_1))
```

Table 2.11: Head of First Test Data frame

grades
80
79
89
74
73
67

5. Students in a statistics class took their first test. The scores they earned are in Table 2.12. Create a density plot for grades. Describe the shape of the distribution. Compare to the graph in question 4.

```
Firsttest_2<- read.csv( "https://krkozak.github.io/MAT160/firsttest_2.csv")
knitr::kable(head(Firsttest_2))
```

Table 2.12: Head of First Test Data frame

grades
67
67
76
47
85
70

2.3 Other Graphical Representations of Data

There are many other types of graphs. Some of the more common ones are the point plot (scatter plot), and a time-series plot. There are also many different graphs that have emerged lately for qualitative data. Many are found in publications and websites. The following is a description of the point plot (scatter plot), and the time-series plot.

2.3.1 Point Plots or Scatter Plot

Sometimes you have two different variables and you want to see if they are related in any way. A scatter plot helps you to see what the relationship would look like. A scatter plot is just a plotting of the ordered pairs.

2.3.2 Example: Scatter Plot

Is there a relationship between systolic blood pressure and weight? To answer this question some data is needed. The data frame NHANES contains this data, but given the size of the data frame, it may be not be very useful to look at the graph of all the data. It makes sense to take a sample from the data frame. A random sample is the better type of sample to take. Once the sample is taken, then a scatter plot can be created. The rStudio command for a scatter plot is

gf_point(response_variable ~ explanatory_variable, data= Data_Frame)

The sample is Table 2.13.

2.3.2.1 Solution

```
sample_NHANES <- NHANES |>
  sample_n(size = 100)
knitr::kable(head(sample_NHANES))
```

ID	SurveyYr	Gender	Age	AgeDecade	AgeMonths	Race1	Race3	Education	MaritalStatus	HHIncome	HH
70839	2011_12	female	13	10-19	NA	White	White	NA	NA	45000-54999	
70653	2011_12	male	43	40-49	NA	White	White	College Grad	NeverMarried	75000-99999	
68507	2011_12	female	49	40-49	NA	Other	Asian	College Grad	NeverMarried	65000-74999	
63084	2011_12	male	4	0-9	NA	White	White	NA	NA	more 99999	
56040	2009_10	female	38	30-39	456	Hispanic	NA	College Grad	LivePartner	more 99999	

	ID	SurveyYr	Gender	Age	AgeDecade	AgeMonths	Race1	Race3	Education	MaritalStatus	HHIncome	HH
62909	2011_12	female	31	30-39		NA	Hispanic	Hispanic	Some College	LivePartner	NA	

Preliminary: State the explanatory variable and the response variable

Let x=explanatory variable = Weight of a person (Weight)

y=response variable = Systolic blood pressure (BPSys1)

```
gf_point(BPSys1~Weight, data=sample_NHANES, xlab="Weight (kg)", ylab="Systolic Blood Pressure", t
```

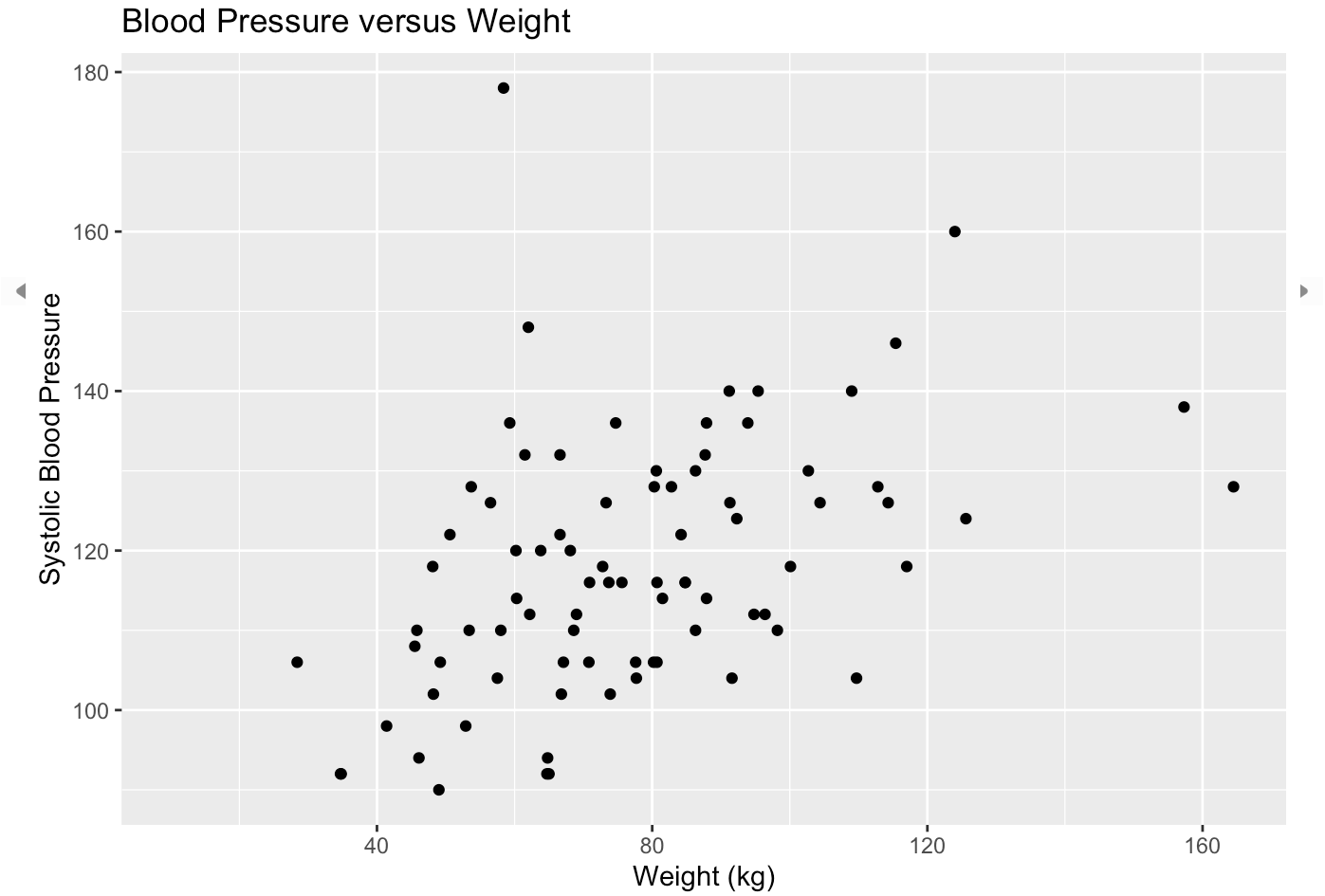


Figure 2.10: Blood Pressure versus Weight

Description of Figure 2.10 is a scatter plot with dots all over the plot though a line could be thought of fitting the dots with lower on the left and higher on the right.

Looking at the graph Figure 2.10, it appears that there is a linear relationship between weight and systolic blood pressure though it looks somewhat weak. It also appears to be a positive relationship, thus as weight increases, the systolic blood pressure increases.

2.3.3 **Time-Series**

A time-series plot is a graph showing the data measurements in chronological order, the data being quantitative data. For example, a time-series plot is used to show profits over the last 5 years. To create a time-series plot on RStudio, use the command

gf_line(response_variable ~ explanatory_variable, data=Data_Frame)

The purpose of a time-series graph is to look for trends over time. Caution, you must realize that the trend may not continue. Just because you see an increase, doesn't mean the increase will continue forever. As an example, prior to 2007, many people noticed that housing prices were increasing. The belief at the time was that housing prices would continue to increase. However, the housing bubble burst in 2007, and many houses lost value, and haven't recovered.

2.3.4 Example: Time-Series Plot

The bank assets (in billions of Australia dollars (AUD)) of the Reserve Bank of Australia (RBA) and other financial organizations for the time period of September 1 1969, through March 1 2019, are contained in table Table 2.14 (Reserve Bank of Australia, 2019). Create a time-series plot of the total assets of Authorized Deposit-taking Institutions (ADIs) and interpret any findings.

```
Australian<- read.csv( "https://krkozak.github.io/MAT160/Australian_financial.csv")
knitr::kable(head(Australian))
```

Date	Day	Assets_RBA	Assets_ADIs_Banks	Assets_ADIs_Building	Assets_ADIs_CU	Assets_ADIs_Total	Assets_RFCs_M
Sep-69	0	2.7	NA	NA	NA	NA	
Dec-69	90	2.9	NA	NA	NA	NA	
Mar-70	180	3.0	NA	NA	NA	NA	
Jun-70	270	3.0	NA	NA	NA	NA	
Sep-70	360	3.0	NA	NA	NA	NA	
Dec-70	450	3.0	NA	NA	NA	NA	

Code book for Data frame Australian

Description The data is a range of economic and financial data produced by the Reserve Bank of Australia and other organizations.

Format

This data frame contains the following columns:

Date: quarters from September 1, 1969, to March 1, 2019

Day: The number of days since September 1, 1969, using 90 days between starts of a quarter. This column is to make it easier to graph in rStudio, and has no other purpose.

Assets_RBA: The assets for the Royal Bank of Australia

Assets_ADIs_Banks: The assets for Authorized Deposit-taking Institutions (ADIs), Banks

Assets_ADIs_Building: The assets for Authorized Deposit-taking Institutions (ADIs), Building societies

Assets_ADIs_CU: The assets for Authorized Deposit-taking Institutions (ADIs), Credit Unions

Assets_ADIs_Total: The assets for Authorized Deposit-taking Institutions (ADIs), total

Assets_RFCs_MM: The assets for Registered Financial Corporations (RFCs), Money Market Corporations

Assets_RFCs_Finance: The assets for Registered Financial Corporations (RFCs), Finance companies and general financiers

Assets_RFCs_Total: The assets for Registered Financial Corporations (RFCs) total

Assets_Life offices: The Assets of Life offices and superannuation funds; Life insurance offices

Assets_Life_funds: The Assets of Life offices and superannuation funds; Superannuation funds

Assets_Life_Total: The Assets of Life offices and superannuation; Total

Assets_Other_Public_trusts: The Assets of Other managed funds; Public unit trusts

Assets_Other_Cash_trusts: The Assets of Other managed funds; Cash management trusts

Assets_Other_Common_funds: The Assets of Other managed funds; Common funds

Assets_Others_Friendly: The Assets of Other managed funds; Friendly societies

Assets_Other_General_insurance: The Assets of Other financial institutions; General insurance offices

Assets_Other_vehicles: The Assets Other financial institutions; Securitisation vehicles

Assets_Unconsolidated: The Assets of Unconsolidated; Statutory funds of life insurance offices; Superannuation

Source Reserve Bank of Australia. (2019, May 13). Statistical Tables. Retrieved July 10, 2019, from https://www.rba.gov.au/statistics/tables/

References Reserve Bank of Australia and other organizations

2.3.4.1 Solution

variable, x=total assets of Authorized Deposit-taking Institutions (ADIs)

Looking at the code book, one can see that the variable Assets_ADIs_Total is the variable in the data frame that is of interest here. With a time series plot, the other variable is time. In this case the variable in the data frame that represents time is Date. The problem with Date is that the units are every quarter. This is not easily interpreted by rStudio, so a column was created called Day. From the code book, this is the number of days since September 1, 1969, using 90 days between starts of a quarter. Even though this isn't perfect, it will work for determining trends. So create a time series plot of Assets_ADIs_Total versus Day. The command is:

```
gf_line(Assets_ADIs_Total~Day, data=Australian, title="Total Assets of Authorized Deposit-taking
```

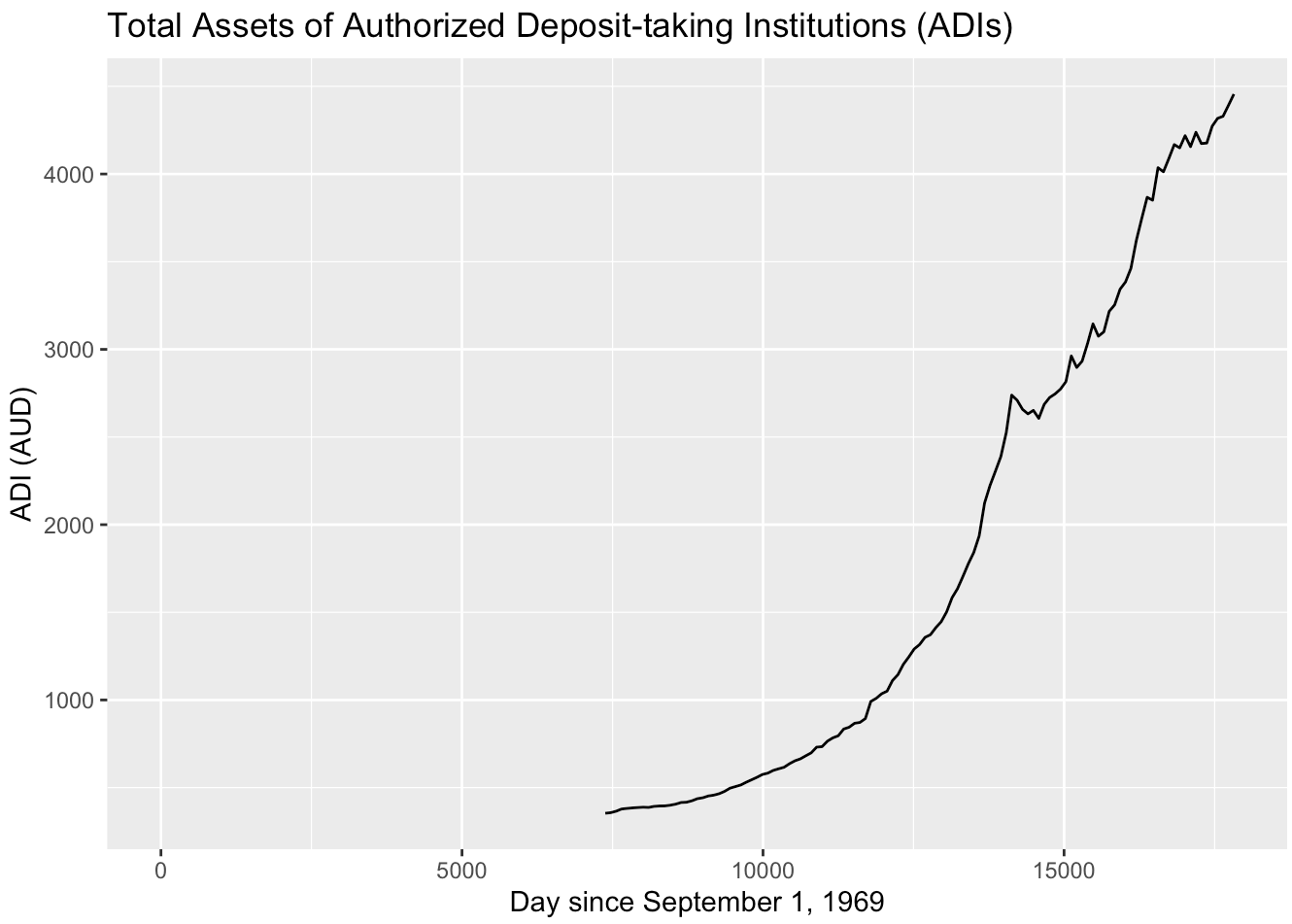


Figure 2.11: Total Assets of Authorized Deposit-taking Institutions

Description of Figure 2.11 is an increasing time series Graph of Total Assets of Authorized Deposit-taking Institutions from day 7500 to 17500. The first number starts at 0 and goes up to about 4500.

From the graph, total assets of Authorized Deposit-taking Institutions (ADIs) appear to be increasing with a slight dip around 14000 days since September 1, 1969. That would be around the year 2008 (14000 days /360 days per year + 1969).

Be careful when making a graph. If the vertical axis doesn't start at 0, then the change can look much more dramatic than it really is. For a graph to be useful to the reader, it needs to have a title that explains what the graph contains, the axes should be labeled so the reader knows what each axes represents, each axes should have a scale marked, and it is best if the vertical axis contains 0 to show the relationship.

2.3.5 Homework for Other Graphical Representations of Data Section

1. When an anthropologist finds skeletal remains, they need to figure out the height of the person. The height of a person (in cm) and the length of one of their metacarpal bone (in cm) were collected and are in Table 2.15 (Prediction of height, 2013). Create a scatter plot of length and height and state if there is a relationship between the height of a person and the length of their metacarpal.

```
Metacarpal<- read.csv( "https://krkozak.github.io/MAT160/metacarpal.csv")
knitr::kable(head(Metacarpal))
```

Table 2.15: Head of Metacarpal Data frame

length	height
45	171
51	178
39	157
41	163
48	172
49	183

Code book for Data frame Metacarpal

Description When anthropologists analyze human skeletal remains, an important piece of information is living stature. Since skeletons are commonly based on statistical methods that utilize measurements on small bones. The following data was presented in a paper in the American Journal of Physical Anthropology to validate one such method.

Format

This data frame contains the following columns:

length: length of Metacarpal I bone in mm

height: stature of skeleton in cm

Source Prediction of Height from Metacarpal Bone Length. (n.d.). Retrieved July 9, 2019, from http://www.statsci.org/data/general/stature.html

References Musgrave, J., and Harneja, N. (1978). The estimation of adult stature from metacarpal bone length. Amer. J. Phys. Anthropology 48, 113-120.

Devore, J., and Peck, R. (1986). Statistics. The Exploration and Analysis of Data. West Publishing, St Paul, Minnesota.

2. The value of the house and the amount of rental income in a year that the house brings in are in Table 2.16 (Capital and rental 2013). Create a scatter plot and state if there is a relationship between the value of the house and the annual rental income.

```
House<- read.csv( "https://krkozak.github.io/MAT160/house.csv")
knitr::kable(head(House))
```

Table 2.16: Head of House Data frame

capital	rental
61500	6656
67500	6864
75000	4992
75000	7280
76000	6656
77000	4576

Code book for Data frame House

Description The data show the capital value and annual rental value of domestic properties in Auckland in 1991.

Format

This data frame contains the following columns:

Capital: Selling price of house in Australian dollar (AUD)

rental: rental price of a house in Australian dollar (AUD)

Source Capital and rental values of Auckland properties. (2013, September 26). Retrieved from http://www.statsci.org/data/oz/rentcap.html

References Lee, A. (1994) Data Analysis: An introduction based on R. Auckland: Department of Statistics, University of Auckland. Data courtesy of Sage Consultants Ltd.

3. The World Bank collects information on the life expectancy of a person in each country (\"Life expectancy at,\" 2013) and the fertility rate per woman in the country (\"Fertility rate,\" 2013). The data for countries for the year 2011 are in Table 2.17. Create a scatter plot of the data and state if there appears to be a relationship between life expectancy and the number of births per woman in 2011.

```
Fertility<- read.csv( "https://krkozak.github.io/MAT160/fertility.csv")
knitr::kable(head(Fertility))
```

Table 2.17: Head of Fertility Data frame

country	lifexp_2011	fertilrate_2011	lifexp_2000	fertilrate_2000	lifexp_1990	fertilrate_1990
Macao SAR, China	79.91	1.03	77.62	0.94	75.28	1.69
Hong Kong SAR, China	83.42	1.20	80.88	1.04	77.38	1.27
Singapore	81.89	1.20	78.05	NA	76.03	1.87
Hungary	74.86	1.23	71.25	1.32	69.32	1.84
Korea, Rep.	80.87	1.24	75.86	1.47	71.29	1.59
Romania	74.51	1.25	71.16	1.31	69.74	1.84

Code book for Data frame Fertility

Description Data is from the World Bank on the life expectancy of countries and the fertility rates in those countries.

Format

This data frame contains the following columns:

Country: Countries in the World

lifexp_2011: Life expectancy of a person born in 2011

fertilrate_2011: Fertility rate in the country in 2011

lifexp_2000: Life expectancy of a person born in 2000

fertilrate_2000: Fertility rate in the country in 2000

lifexp_1990: Life expectancy of a person born in 1990

fertilrate_1990: Fertility rate in the country in 1990

Source Life expectancy at birth. (2013, October 14). Retrieved from http://data.worldbank.org/indicator/SP.DYN.LE00.IN

References Data from World Bank, Life expectancy at birth, total (years)

4. The World Bank collected data on the percentage of gross domestic product (GDP) that a country spends on health expenditures (Current health expenditure (% of GDP), 2019), the fertility rate of the country (Fertility rate, total (births per woman), 2019), and the percentage of women receiving prenatal care (Pregnant women receiving prenatal care (%), 2019). The data for the countries where this information is available in Table 2.18. Create a scatter plot of the health expenditure and percentage of women receiving prenatal care in the year 2014, and state if there appears to be a relationship between percentage spent on health expenditure and the percentage of women receiving prenatal care.

```
Fert_prenatal<-read.csv( "https://krkozak.github.io/MAT160/fertility_prenatal.csv")
knitr::kable(head(Fert_prenatal))
```

Country.Name	Country.Code	Region	IncomeGroup	f1960	f1961	f1962	f1963	f1964	f1965	f1966	f1967	f19
Angola	AGO	Sub-Saharan Africa	Lower middle income	7.478	7.524	7.563	7.592	7.611	7.619	7.618	7.613	7.6
Armenia	ARM	Europe & Central Asia	Upper middle income	4.786	4.670	4.521	4.345	4.150	3.950	3.758	3.582	3.4
Belize	BLZ	Latin America & Caribbean	Upper middle income	6.500	6.480	6.460	6.440	6.420	6.400	6.379	6.358	6.3
Cote d'Ivoire	CIV	Sub-Saharan Africa	Lower middle income	7.691	7.720	7.750	7.781	7.811	7.841	7.868	7.893	7.9
Ethiopia	ETH	Sub-Saharan Africa	Low income	6.880	6.877	6.875	6.872	6.867	6.864	6.867	6.880	6.9
Guinea	GIN	Sub-Saharan Africa	Low income	6.114	6.127	6.138	6.147	6.154	6.160	6.168	6.177	6.1

Code book for Data frame Fert_prenatal

Description Data is from the World Bank on money spent on expenditure of countries and the percentage of women receiving prenatal care in those countries.

Format

This data frame contains the following columns:

Country.Name: Countries around the world

Country.Code: Three letter country code for countries around the world

Region: Location of a country around the world as classified by the World Bank

IncomeGroup: The income level of a country as classified by the World Bank

f1960-f2017: Fertility rate of a country from 1960-2017

p1986-p2018: Percentage of women receiving prenatal care in the country in 1986-2018

e200-2016: Expenditure amounts of the countries for medical care in 2000-2016 (% of GDP)

Source Fertility rate, total (births per woman). (n.d.). Retrieved July 8, 2019, from https://data.worldbank.org/indicator/SP.DYN.TFRT.IN Pregnant women receiving prenatal care (%). (n.d.). Retrieved July 9, 2019, from https://data.worldbank.org/indicator/SH.STA.ANVC.ZS Current health expenditure (% of GDP). (n.d.). Retrieved July 9, 2019, from https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS

References Data from World Bank, fertility rate, expenditure on health, and pregnant woman rate of prenatal care.

5. The Australian Institute of Criminology gathered data on the number of deaths (per 100,000 people) due to firearms during the period 1983 to 1997 (\"Deaths from firearms,\" 2013). The data is in Table 2.19. Create a time-series plot of the data and state any findings you can from the graph.

```
Firearm<- read.csv( "https://krkozak.github.io/MAT160/rate.csv")
knitr::kable(head(Firearm))
```

Table 2.19: Head of Firearm Data frame

year	rate
1983	4.31
1984	4.42
1985	4.52
1986	4.35
1987	4.39
1988	4.21

Code book for Data Frame Firearm

Description The data give the number of deaths caused by firearms in Australia from 1983 to 1997, expressed as a rate per 100,000 of population.

Format

This data frame contains the following columns:

Year: Years from 1983 to 1997

Rate: Rate of deaths caused by firearms in Australia per 100,000 population

Source Deaths from firearms. (2013, September 26). Retrieved from http://www.statsci.org/data/oz/firearms.html

References Australian Institute of Criminology, 1999.The data was contributed by Rex Boggs, Glenmore State High School, Rockhampton, Queensland, Australia.

6. The economic crisis of 2008 affected many countries, though some more than others. Some people in Australia have claimed that Australia wasn't hurt that badly from the crisis. The bank assets (in billions of Australia dollars (AUD)) of the Reserve Bank of Australia (RBA) for the time period of September 1 1969, through March 1 2019, are contained in @bl-Australian (Reserve Bank of Australia, 2019). Create a time-series plot of the assets of the RBA and interpret any findings.

Code book for Data Frame Australian is below Table 2.14.

7. The consumer price index (CPI) is a measure used by the U.S. government to describe the cost of living. The cost of living for the U.S. from the years 1913 through 2019, with the year 1982 being used as the year that all others are compared (Consumer Price Index Data from 1913 to 2019, 2019) is given in Table 2.20. Create a time-series plot of the Average Annual CPI for 2018 and interpret.

```
CPI<- read.csv( "https://krkozak.github.io/MAT160/CPI_US.csv")
knitr::kable(head(CPI))
```

Table 2.20: Head of CPI Data frame

Year	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec	Annual_avg	PerDec_Dec	Perc_Avg_Avg
1913	9.8	9.8	9.8	9.8	9.7	9.8	9.9	9.9	10.0	10.0	10.1	10.0	9.9	–	–
1914	10.0	9.9	9.9	9.8	9.9	9.9	10.0	10.2	10.2	10.1	10.2	10.1	10.0	1	1
1915	10.1	10.0	9.9	10.0	10.1	10.1	10.1	10.1	10.1	10.2	10.3	10.3	10.1	2	1
1916	10.4	10.4	10.5	10.6	10.7	10.8	10.8	10.9	11.1	11.3	11.5	11.6	10.9	12.6	7.9
1917	11.7	12.0	12.0	12.6	12.8	13.0	12.8	13.0	13.3	13.5	13.5	13.7	12.8	18.1	17.4
1918	14.0	14.1	14.0	14.2	14.5	14.7	15.1	15.4	15.7	16.0	16.3	16.5	15.1	20.4	18

Code book for Data frame CPI

Description This table of Consumer Price Index (CPI) data is based upon a 1982 base of 100.

Format

This data frame contains the following columns:

Year: Year from 1913 to 2019

Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec: CPI for a particular month

Average_Avg: The average CPI for a particular year

PerDec_Dec: Percent change from December to December

Per_Avg_Avg: Percent change from Annual Average to Annual Average

Source Consumer Price Index Data from 1913 to 2019. (2019, June 12). Retrieved July 10, 2019, from https://www.usinflationcalculator.com/inflation/consumer-price-index-and-annual-percent-changes-from-1913-to-2008/

References US Inflation Calculator website, 2019.

8. The mean and median incomes income in current dollars is given in Table 2.21. Create a time-series plot and interpret.

```
US_income<- read.csv( "https://krkozak.github.io/MAT160/US_income.csv")
knitr::kable(head(US_income))
```

Table 2.21: Head of US_income Data frame

year	number	med_income_current	med_income_2017	mean_income_current	mean_income_2017
2017	127586	61372	61372	86220	86220
2016	126224	59039	60309	83143	84931
2015	125819	56516	58476	79263	82012
2014	124587	53657	55613	75738	78500
2013	122952	51939	54744	72641	76565
2012	122459	51017	54569	71274	76237

Code book for Data Frame US_income

Description This table is of US mean and median incomes in both current dollars and in 2017 dollars.

Format

This data frame contains the following columns:

Year: Year from 1975 to 2017

number: Households as of March of the following year. (in thousands)

med_income_current: median income of a US household in current dollars

med_income_2017: median income of a US household in 2017 CPI-U-RS adjusted dollars

mean_income_current: mean income of a US household in current dollars

mean_income_2017: mean income of a US household in 2017 CPI-U-RS adjusted dollars

Source US Census Bureau. (2018, March 06). Data. Retrieved July 21, 2019, from https://www.census.gov/programs-surveys/cps/data-detail.html

References U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplements.