

Chapter 1

Descriptive Statistics

In this chapter we will discuss the following topics:

- How to make Pie charts with the R-function **pie()**.
- How to make Barplots with the R-function **barplot()**.
- How to make Histograms with the R-function **hist()**.
- How to make Stemplots with the R-function **stem()**.

Pie Charts

A Pie chart displays the distribution of a categorical variable into slices such that the size of each slice represents the counts or proportion of each category. All the categories must be included in the pie chart and each category is related to the whole.

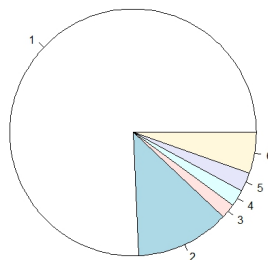
Problem. The table below shows data from 2007-2011 on the percents of types of transportation to work in Arizona [1].

	Transportation Percentage
Car, truck, or van - drove alone	75.76
Car, truck, or van - carpooled	12.25
Public transportation (excluding taxicab)	2.00
Walked	2.15
Taxicab, motorcycle, bicycle, or other means	2.49
Worked at home	5.35

Create a simple pie chart.

Solution.

```
> Transportation=c(75.76, 12.25, 2.00, 2.15, 2.49, 5.35)
> pie(Transportation)
```



Problem. Create the pie chart shown in the picture below.

Solution.

```
> Transportation=c(75.76, 12.25, 2.00, 2.15, 2.49, 5.35)
> colors=c("gray90","darkgreen","grey70","blue","yellow","brown")
> pie(Transportation, main="Means of Transportation to Work in Arizona",
+ col=colors, labels=Transportation, cex=0.7)
> legend("topright", c("Car, truck, or van - drove alone", "truck, or van - carpooled",
+ "Public transportation (excluding taxicab)","Walked",
+ "Taxicab, motorcycle, bicycle, or other means","Worked at home"),
+ cex=0.5,fill=colors)
```

Means of Transportation to Work in Arizona

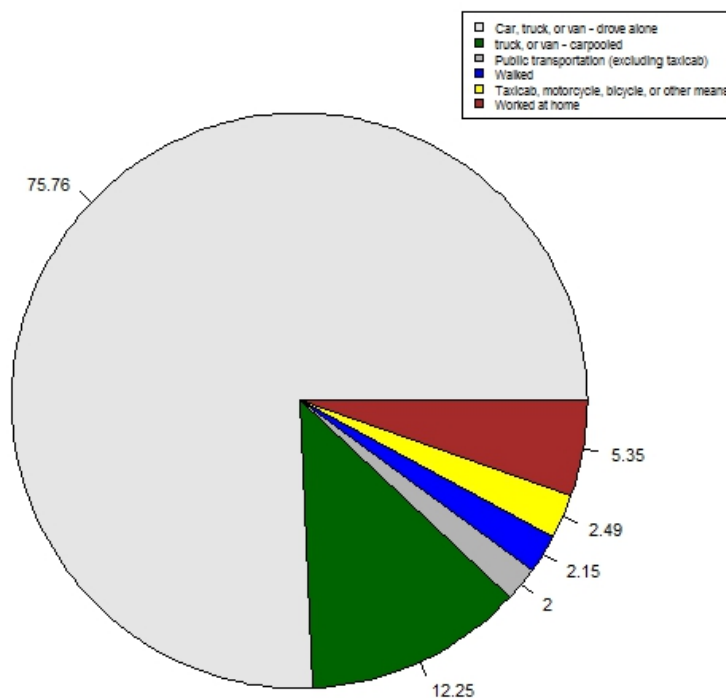


Figure 1: Source: U.S. Census Bureau, 2007-2011 American Community Survey 5-Year Estimates.

Explanation. The explanation of these steps are as follows:

- The command `c()` creates an array of numbers and the command `Transportation=c()` creates a vector variable named **Transportation**.

- The command

```
colors=c("gray90","darkgreen","grey70","blue","yellow","brown")
```

creates an array of text strings whose elements are given in quotes.

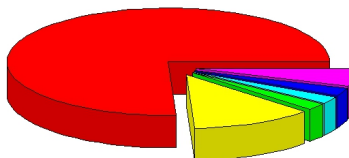
- Adding the entry `main=""` in `pie()` gives a title to the pie chart.
- Adding the entry `col` sets the colors of the sectors.
- The entry `labels` specifies the percentages to each sector. It must carry the same name as the first entry in the `pie()` command.
- The entry `cex=0.7` is a scaling factor for the font size in the pie relative to the font size of the main title.
- The command `legend()` creates the legend, and the entry `"topright"` places the legend in the upper right corner of the plot.
- The entry `cex=0.5` is a scaling factor for the font size in the legend relative to the font size of the main title.
- The entry `fill=colors` fills in the colors in the legend.

Problem. (For more advanced R-users:) Plot the pie chart in 3 dimensions.

Solution.

```
> library(plotrix)
> Transportation=c(75.76, 12.25, 2.00, 2.15, 2.49, 5.35)
> pie3D(Transportation, explode=0.1, main="Means of Transportation to Work in Arizona")
```

Means of Transportation to Work in Arizona



Explanation. The code can be explained as follows:

- The command `library(plotrix)` calls for the library named `plotrix`. This library must be installed in R before it can be used.
- The command `pie3D` makes a 3-D pie chart.
- The entry `explode=0.1` is a factor that determines the amount of detachment of the slices in the pie.

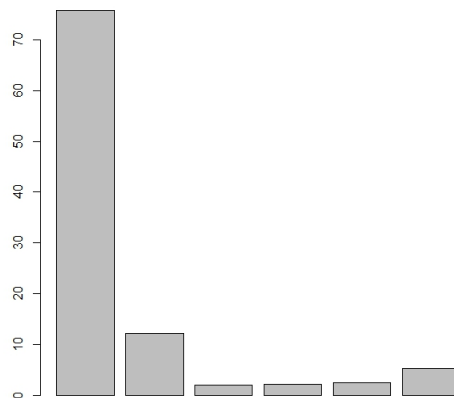
Barplots

A Barplot is a graphical display of the distribution of a categorical variable. Each bar represent a category and the height of the bars shows the counts or percents for the categories.

Problem. Create a simple barplot.

Solution.

```
> Transportation=c(75.76, 12.25, 2.00, 2.15, 2.49, 5.35)
> barplot(Transportation)
```



Problem. Create a barplot that includes the legend.

Solution.

```
> Transportation=c(75.76, 12.25, 2.00, 2.15, 2.49, 5.35)
> colors=c("gray90", "darkgreen", "grey70", "blue", "yellow", "brown")
> Transportationarray=array(c(Transportation), dim=c(6,1))
> barplot(Transportationarray, main="Means of Transportation to Work in Arizona",
+ ylab= "percent", beside=TRUE, names.arg=c("Transportation"),col=colors)
> legend("topright", c("Car, truck, or van - drove alone", "truck, or van - carpooled",
+ "Public transportation (excluding taxicab)", "Walked",
+ "Taxicab, motorcycle, bicycle, or other means", "Worked at home"),
+ cex=0.8,fill=colors)
```

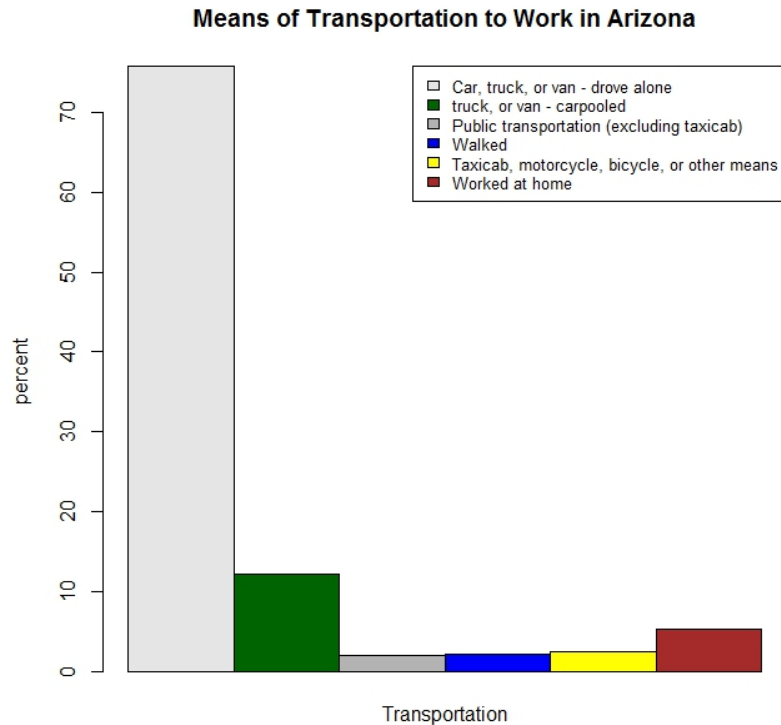


Figure 2: Source: U.S. Census Bureau, 2007-2011 American Community Survey 5-Year Estimates.

Explanation. The code can be explained as follows:

- The command `array(c(Transportation), dim=c(6,1))` creates a matrix with 6 rows and 1 column.
- Adding the entry `main=""` gives a title.
- Adding the entry `ylab=""` labels the y-axis.
- Adding the entry `beside=TRUE` places the bars next to each other.
- Adding the entry `names.arg=c("Transportation")` labels the base of the bars as Transportation.
- Adding the entry `col` sets the colors of the bars.
- The command `legend()` is explained above under the section about Pie charts.

Histograms

A histogram is a graphical display of the distribution of a quantitative variable.

Problem. The following table shows the average point per game by players for the University of Arizona's men's basketball team for the 2012-2013 season. [2]

	Player	PPG
1	Mark Lyons	15.6
2	Solomon Hill	13.4
3	Nick Johnson	11.5
4	Kevin Parrom	8.3
5	Brandon Ashley	7.5
6	Kaleb Tarczewski	6.6
7	Grant Jerrett	5.2
8	Jordin Mayes	2.8
9	Gabe York	2.4
10	Angelo Chol	1.9
11	Max Wiepking	1.1
12	Quinton Crawford	0.0
13	Drew Mellon	0.0
14	Jacob Hazzard	0.0

Create a simple histogram and let R specify the bin.

Solution.

```
> PPG=c(15.6, 13.4, 11.5, 8.3, 7.5, 6.6, 5.2, 2.8, 2.4, 1.9, 1.1, 0.0, 0.0, 0.0)
> hist(PPG,col="purple")
```

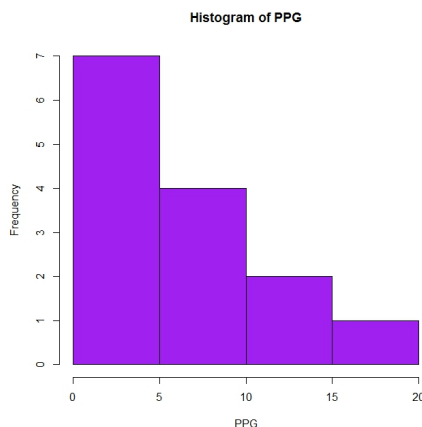


Figure 3: Source: Arizona wildcat statistics 2012-2013

This type of histogram is called a *frequency histogram*. The base is the class length and the height of each rectangle represents the frequency or number of observations in each class/group.

Problem. If we add the argument **prob=TRUE** in R, we will create a *relative frequency histogram* for which the area of each rectangle is the relative frequency of the observations in each class.

Problem. Create a *relative frequency histogram*.

Solution. We obtain:

```
> PPG=c(15.6, 13.4, 11.5, 8.3, 7.5, 6.6, 5.2, 2.8, 2.4, 1.9, 1.1, 0.0, 0.0, 0.0)
> hist(PPG,col="yellow",prob=TRUE)
```

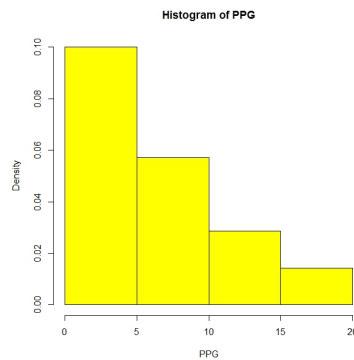


Figure 4: Source: Arizona wildcat statistics 2012-2013

Create a histogram and specify the cutpoints.

Solution. We let the cutpoints be located at $-0.2, 3.8, 7.8, 11.8, 15.8$. We obtain:

```
> PPG=c(15.6, 13.4, 11.5, 8.3, 7.5, 6.6, 5.2, 2.8, 2.4, 1.9, 1.1, 0.0, 0.0, 0.0)
> hist(PPG,breaks=seq(-0.2,15.8,4),xlab="Average point per game",
+ main="Points per game for U of A, Men's basketball team in 2012-2013",col="green")
```

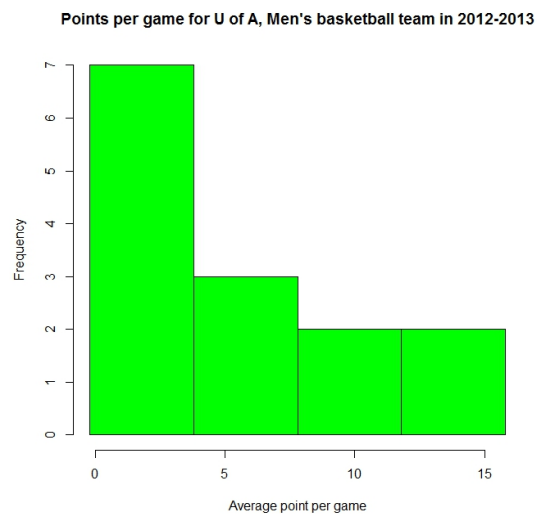


Figure 5: Source: Arizona wildcat statistics 2012-2013

Explanation. The code can be explained as follows:

- The entry `seq(-0.2,15.8,4)` generates an array of equally spaced points from -0.2 to 15.8 in steps of 4 .
- The entry `breaks=seq(-0.2,15.8,4)` specifies the cutpoints. The cutpoints are here $-0.2, 3.8, 7.8, 11.8, 15.8$, and the bin length is 4 .
- Adding the entry `xlab=""` labels the x-axis.
- Adding the entry `main=""` gives a title.

Stemplots

A stemplot is another graphical display of the distribution of a quantitative variable. In R the observations are ordered from the smallest to the largest in each leaf.

Problem. Draw a stemplot.

Solution.

```
> PPG=c(15.6, 13.4, 11.5, 8.3, 7.5, 6.6, 5.2, 2.8, 2.4, 1.9, 1.1, 0.0, 0.0, 0.0)
> stem(PPG)
```

The decimal point is 1 digit(s) to the right of the |

```
0 | 0001223
0 | 5788
1 | 23
1 | 6
```

Problem. Draw a stemplot that includes more stems.

Solution.

```
> stem(PPG,scale=3)
```

The decimal point is at the |

```
0 | 00019
2 | 48
4 | 2
6 | 65
8 | 3
10 | 5
12 | 4
14 | 6
```

References

- [1] U.S. Census Bureau, 2007-2011 American Community Survey 5-Year Estimates at <http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml>

[2] Arizona wildcat statistics 2012-2013 at
http://espn.go.com/mens-college-basketball/team/stats/_/id/12/year/2013/arizona-wildcats

00