

Intro to RMarkdown

Katie Lankowicz

2023-06-26

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

To make this work, you need to download a separate software called MikTeX. Please follow the instructions here: <https://webpages.uidaho.edu/~renaes/419/Handouts/R%20markdown%20prep.pdf>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean   : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.   :120.00
```

Look very closely at the way the code chunk is written in the text editor. Everything is important here. You need three *accent* signs, the squiggly brackets, and *r* written to begin a code chunk, then you end the code chunk with a second set of *accent* signs. The code chunk also has the word *cars*, this is simply the name of the chunk. You should name chunks something that indicates their purpose.

If you want a code chunk to accomplish a task but DO NOT want the actual code to be repeated in the compiled report, you can use the following syntax:

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean   : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.   :120.00
```

If you wrote a code chunk that you do not think you actually need and do not need to run, but want to keep just in case, use the following syntax to suppress both the evaluation of the chunk:

```
summary(cars)
```

You can use both *echo* and *eval* for a single chunk to essentially make it disappear from all output.

Loading data

Loading data occurs the same way in RMarkdown as it does for regular R scripts. We will be working with data from the Behavioral Risk Factor Surveillance System (BRFSS), which is an annual telephone survey of 350,000 people in the United States. Respondents are asked about their diet and physical activity, HIV/AIDS status, tobacco use, and healthcare coverage. We will use a random sample of 20,000 people pulled from this dataset. I have saved this as a CSV file in our GitHub repository, which we can download at will.

```
setwd(here("Learning_R"))
cdc <- read.csv("cdc.csv")
```

Summaries

Let's go over some of the concepts we learned about last week, starting with a brief summary of the data.

```
##      genhlth  exerany  hlthplan  smoke100  height  weight  wt desire  age  gender
## 1      good      0         1         0      70     175     175  77      m
## 2      good      0         1         1      64     125     115  33      f
## 3      good      1         1         1      60     105     105  49      f
## 4      good      1         1         0      66     132     124  42      f
## 5 very good      0         1         0      61     150     130  55      f
## 6 very good      1         1         0      64     114     114  55      f

## [1] 169.683

## [1] 165

## [1] 1606.484
```

Tables

We can determine the number of respondents who smoke using a table, like this. In this dataset, 0 represents non-smokers and 1 represents smokers.

```
table(cdc$smoke100)
```

```
##  
##      0      1  
## 10559  9441
```

Let's say we want to look at the proportion of smokers to non-smokers. This would be equal to dividing the number of smokers by the population size (20,000 respondents) and the number of non-smokers by the population size.

```
table(cdc$smoke100) / nrow(cdc)
```

```
##  
##      0      1  
## 0.52795 0.47205
```

If we want to show the breakdown of smokers by gender, we can make a two-way table.

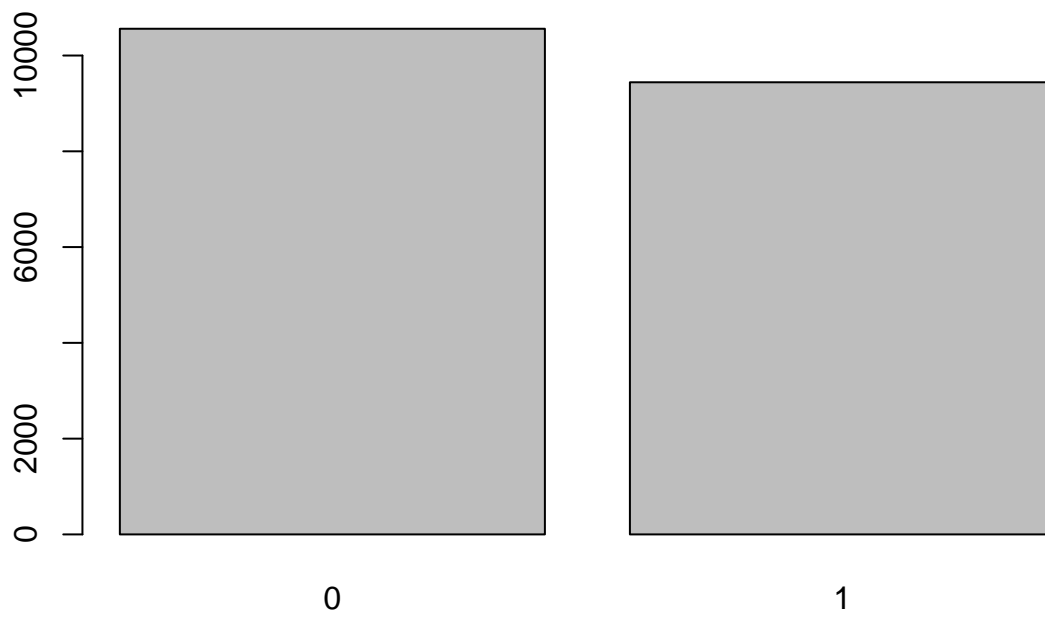
```
table(cdc$gender, cdc$smoke100)
```

```
##  
##      0      1  
## f 6012 4419  
## m 4547 5022
```

Plots

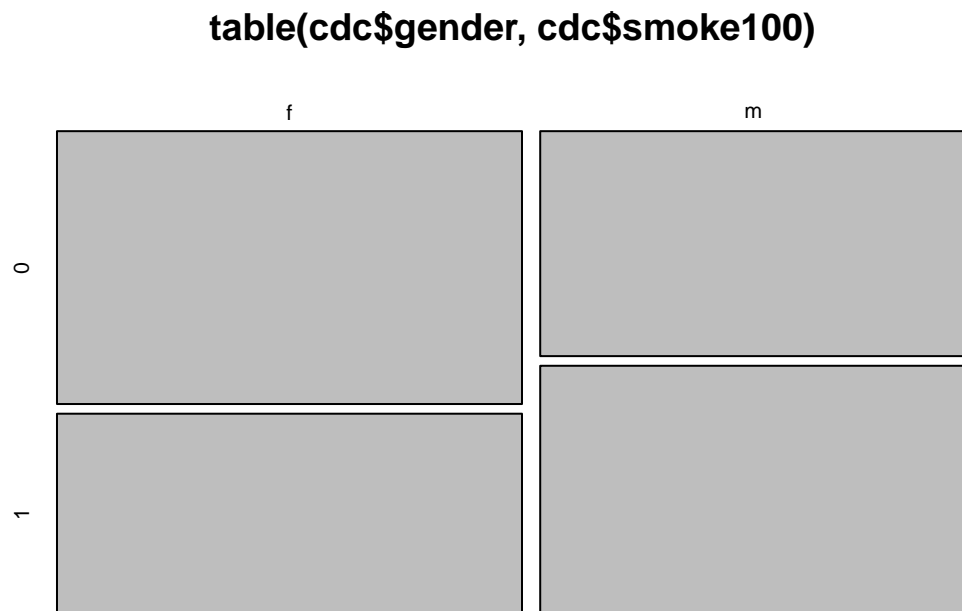
We can turn this information into a barplot. RMarkdown is a great tool for data visualization. It easily embeds plots into your output.

```
barplot(table(cdc$smoke100))
```



Let's visualize the relationship between gender and smoking. This would make for a great mosaic plot, in which we can check the relative frequency of a dependent variable (smoking) among different categorical variables (gender)

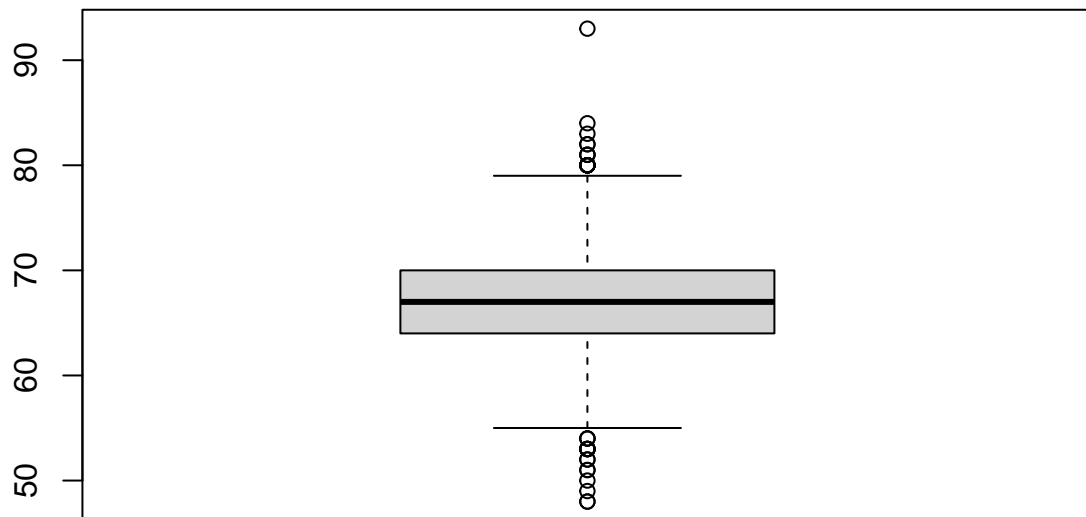
```
mosaicplot(table(cdc$gender, cdc$smoke100))
```



Quantitative data

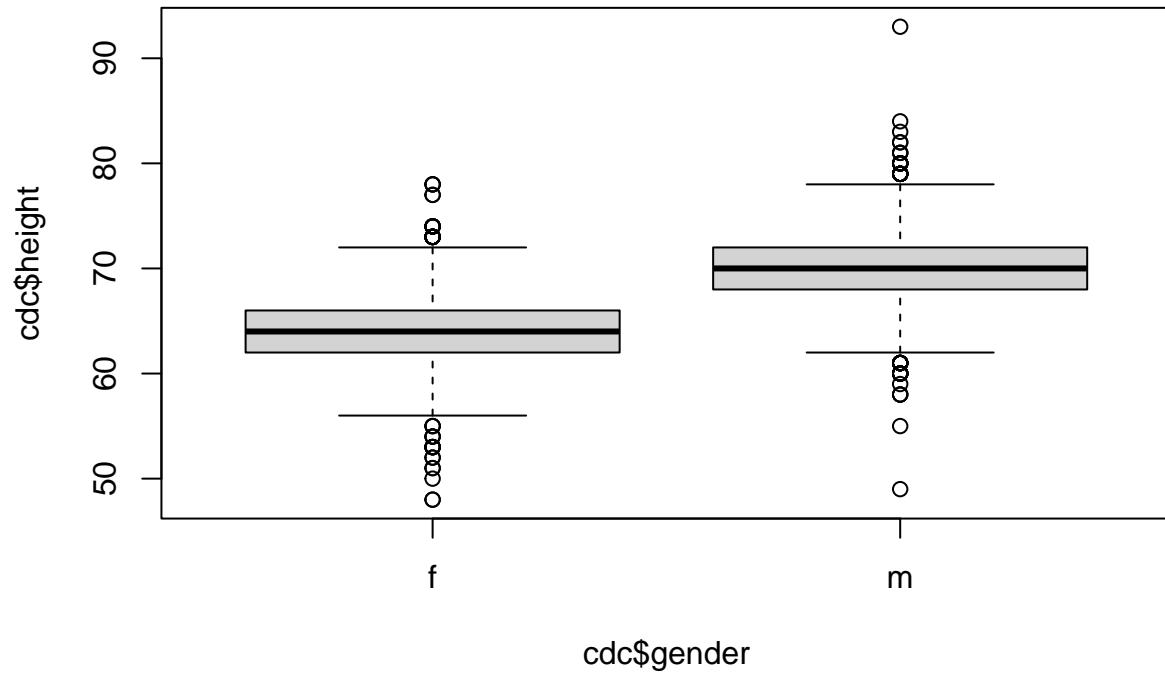
So far, we have worked on categorical data (gender, yes or no to smoking). Let's dig in to some quantitative stuff. We can make box-and-whisker plots using the quantitative data of respondent height.

```
boxplot(cdc$height)
```



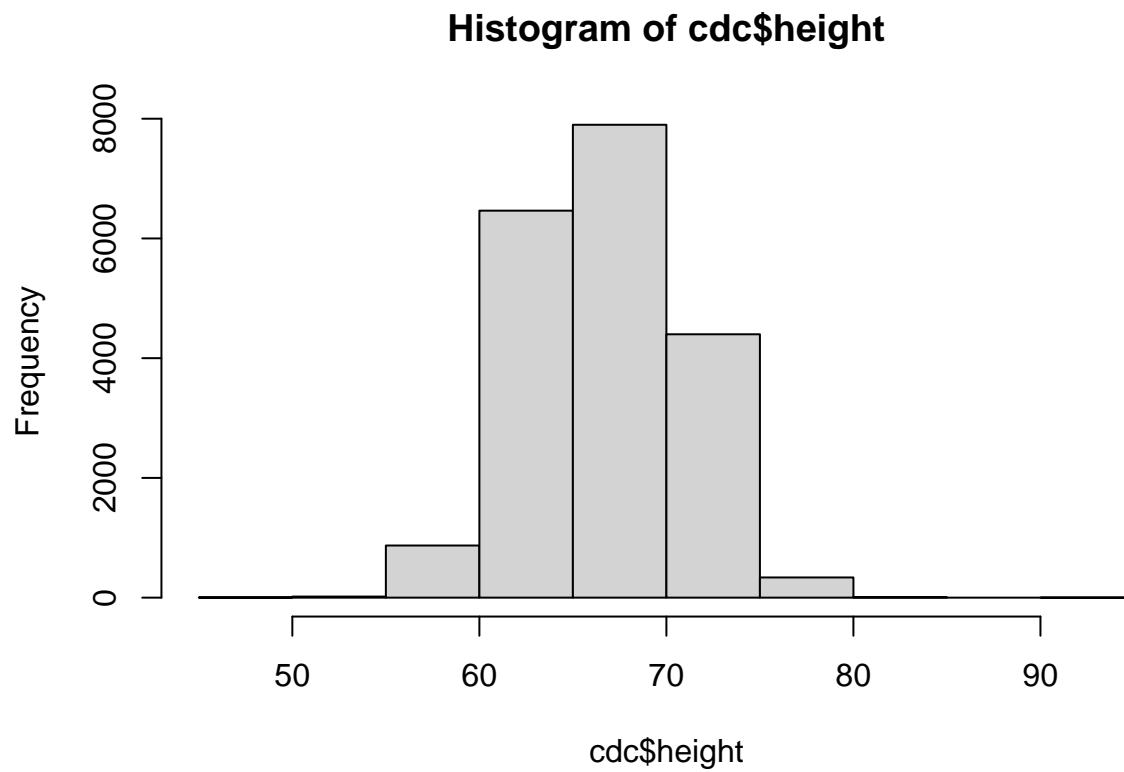
We can break this into categories, as well.

```
boxplot(cdc$height ~ cdc$gender)
```



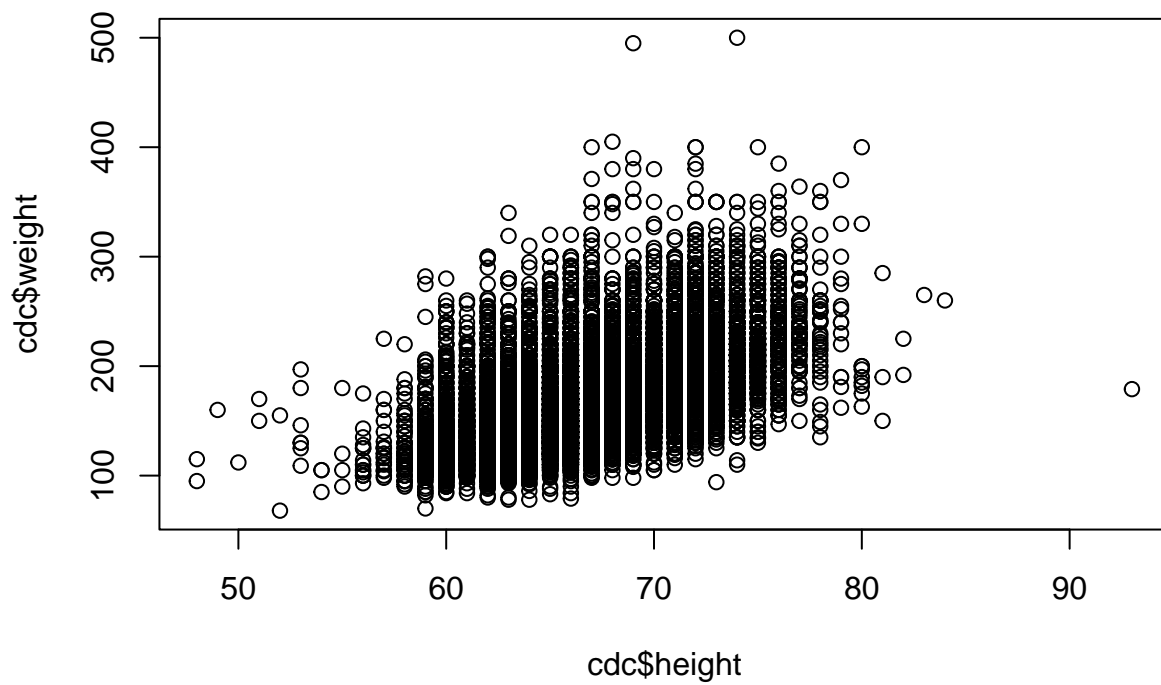
We can also make a histogram of respondent height

```
hist(cdc$height)
```



Or a scatterplot of respondent height vs. weight

```
plot(cdc$height, cdc$weight)
```



Saving and getting output

Once you're done editing the markdown document, you can save it normally. You can also “knit” it to get a nice document showing all your work and output. Click the “Knit” button above, and you'll see the “Render” tab in the console become active. RStudio will then work with MikTeX to create whatever kind of document you have specified (html, pdf, or microsoft word). You can then navigate to your working directory and view the output.