

STAT 1261/2260: Principles of Data Science

Final Project Part I

Caitlin Alano

10/18/2020

Introduction: Installing and Loading fivethirtyeightdata package

```
# installed fivethirtyeightdata package using code below
# install.packages('fivethirtyeightdata', repos = 'https://fivethirtyeightdata.github.io/drat/', type = 'source')

# load steak_survey dataset
library(fivethirtyeight)
library(fivethirtyeightdata)
steak_survey
```

Recreating Graphic - Bar Graph/Plot

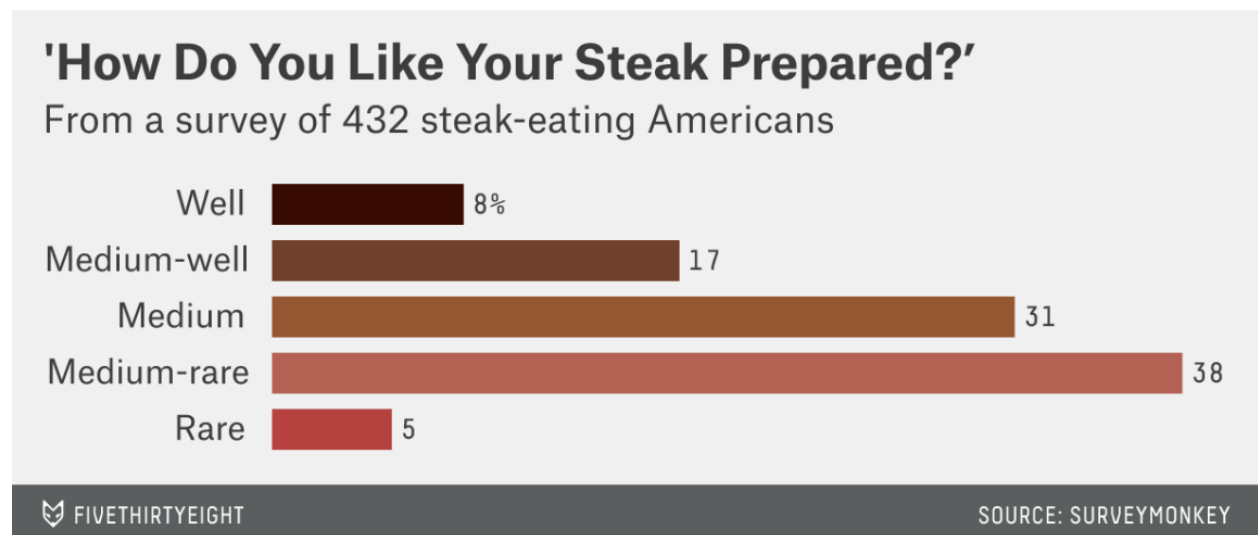


Figure 1: ORIGINAL graphic

```
# load all necessary packages
library(ggplot2)
library(fivethirtyeight)
library(fivethirtyeightdata)
library(mdsr)
```

```
##
## Attaching package: 'mdsr'
```

```
## The following object is masked from 'package:utils':
##
##   person

library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --

## v tibble  3.0.3      v dplyr    1.0.2
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0
## v purrr   0.3.4

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

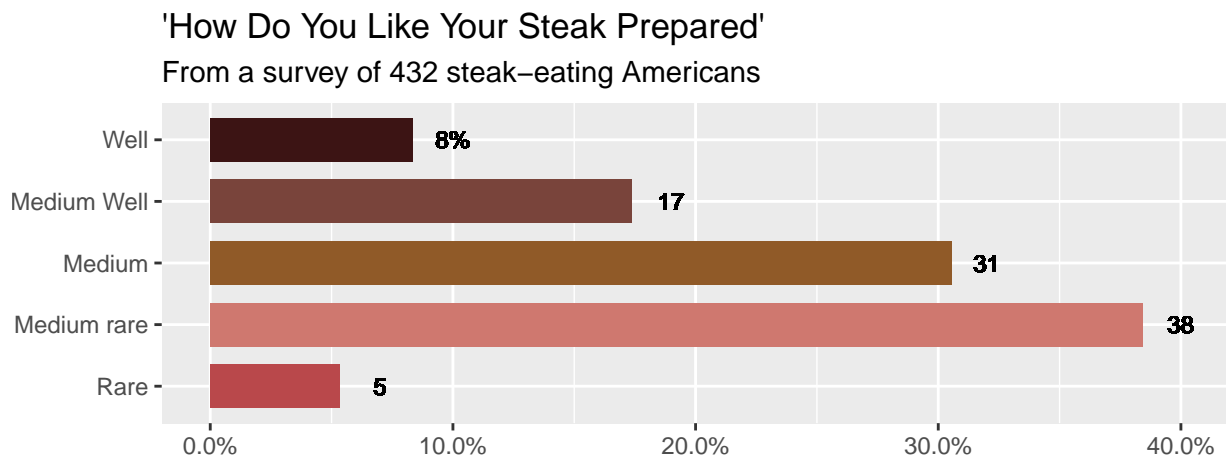
library(dplyr)
library(RColorBrewer)

cols <- c("Rare" = "#b9484b", "Medium rare" = "#cf786f", "Medium" = "#905a28", "Medium Well" = "#79443b", "Well" = "#4d2c2c")

steak_plot <- ggplot(data = steak_survey, mapping = aes(x = steak_prep, y = ..prop.., group=1)) +
  geom_bar(mapping = aes(x = steak_prep, y = ..prop.., group = 1, na.rm = TRUE,), width = 0.7, fill = cols) +
  scale_fill_manual(values = cols) +
  theme(aspect.ratio = .3) +
  scale_x_discrete(na.translate = FALSE) +
  scale_y_continuous(labels = scales::percent) +
  labs(x = NULL, y = NULL) +
  ggtitle("'How Do You Like Your Steak Prepared'", subtitle = "From a survey of 432 steak-eating Americans") +
  geom_text(aes(y=0.07,x=1),label="5",cex=3)+
  geom_text(aes(y=0.40,x=2),label="38",cex=3)+
  geom_text(aes(y=0.32,x=3),label="31",cex=3)+
  geom_text(aes(y=0.19,x=4),label="17",cex=3)+
  geom_text(aes(y=0.10,x=5),label="8%",cex=3)+
  coord_flip()

## Warning: Ignoring unknown aesthetics: na.rm
steak_plot
```

```
## Warning: Removed 118 rows containing non-finite values (stat_count).
```



I. Data Wrangling-Visualization Statements Technical Discussion

First, I made a name 'steak_plot' for the bar graph to recreate. Then, using ggplot2, I used the ggplot() function and specified the data being used from the fivethirtyeightdata called 'steak_survey' in this function. Then, in the same function, I used aesthetic mapping, including the x-variable which would be how people wanted their steak prepared through the variable 'steak_prep,' as well as the y variable, which we wanted to be proportion, so I used '.prop.'. I then included group = 1 as a "dummy" grouping in order to override default behavior where geom_bar groups the x variable in order to separately count rows per level of the x variable. When each level is considered separately, the proportion of each preparation level in itself is always going to be 100%, making all bars equally as tall at 100%. Instead, since I wanted accurate proportions calculated, I need all levels of 'steak_prep' to be considered together. group = 1 makes sure that proportions of each level of 'steak_prep' will be relative to **all** levels of 'steak_prep'.

The next issue I ran into was making sure 'NA' values for 'steak_prep' were not included in the graph, as originally they would appear in their own bar. I fixed this by using the scale_x_discrete() function with 'na.translate=FALSE' argument. Since this discrete scale easily showed missing values by default, I used this function to remove the NA values. I also had to alter the y-axis increment labels to change them to percentages by using the scale_y_continuous() function with the argument 'labels = scales::percent'.

Since the graph did not need labels for the x and y axes, I used the labs() function with arguments 'x = NULL' and 'y = NULL', but I included the title and subtitle from the original graphic using the ggtitle() function containing both as arguments.

Next, I needed to have the labels for percentages on the right side of each bar (in terms of after I flipped the coordinate system, which I will explain later.) In order to do this, I used the geom_text() function, and added the necessary arguments including aes(y=, x=) where I used values for position, label= where I used values for the percentages of each steak preparation category, and cex= which indicated the amount by which plotting text and symbols should be scaled relative to the default(1).

Afterwards, I assigned colors to each category of 'steak_prep'. First, I found colors that looked very similar to the colors in the original graphic and found their hex codes. Then I made a vector called 'cols' where I assigned the hex codes to their respective category in 'steak_prep'. In order to apply this to the bar plot, I used the 'fill=' argument in the geom_bar() function and set it equal to the 'cols' vector. Afterwards, I could use the scale_fill_manual() with the 'values=' argument, which I set equal to 'cols' as well.

Then, to reduce the width of the bars, I added the 'width=' argument in the geom_bar() function. Afterwards, to reduce the space in between bars, I used the theme() function with the 'aspect.ratio=' argument.

Since I needed flip the x and y axes, I used the coord_flip() function. Finally, I printed out the bar plot by adding its name 'steak_plot' at the end.

II. Data Graphic Context

This data analysis originally started off with writer Walt Hickey's goal of determining whether or not risk-averse people are more likely to order their steak well-done. However, he included this graphic instead, noting that it is the only relevant outcome of this analysis. The graphic itself displays how respondents prefer their steak, from a survey of 432 Americans by calculating percentages per level of steak preparation, with the majority preferring their steak to be prepared either medium or medium-rare with a combined total of 69%.

III. Title and Hyperlink to Original Article

Title: How Americans Like Their Steak

<https://fivethirtyeight.com/features/how-americans-like-their-steak/>