SipAndCode

Eleanor Chodroff

7/2/2018

# R-Ladies Chicago Sip & Code  
# Code from our collaborative coding session analyzing the wine ratings collected at the event  
# This served as an introduction to R coding for many group members and was a first pass analysis through the data.  
  
# This code is organized \*chronologically\* as it was primarily used for teaching purposes  
  
# import dataset using Import Dataset... option and readr package in RStudio  
library(readr)  
wine <- read\_csv("~/Downloads/2018-07-02-wine-tasting-master/Wine Tasting\_RLadies.csv")

## Parsed with column specification:  
## cols(  
## Wine = col\_character(),  
## Score = col\_integer()  
## )

# preview the dataset in RStudio  
View(wine)  
  
# examine the grand average wine rating  
mean(wine$Score)

## [1] 86.3

# rename the file because one of the column names is highly similar to the dataset name  
# this can sometimes cause headaches that are best avoided  
ratings <- wine  
  
# remove old dataset  
rm(wine)  
  
# preview the dataset  
View(ratings)  
  
# take a look at overall spread of ratings  
# this function returns a summary of each column in the dataset  
summary(ratings)

## Wine Score   
## Length:40 Min. :69.0   
## Class :character 1st Qu.:83.0   
## Mode :character Median :85.5   
## Mean :86.3   
## 3rd Qu.:90.0   
## Max. :97.0

# just look at the summary for the ratings$Score column  
summary(ratings$Score)

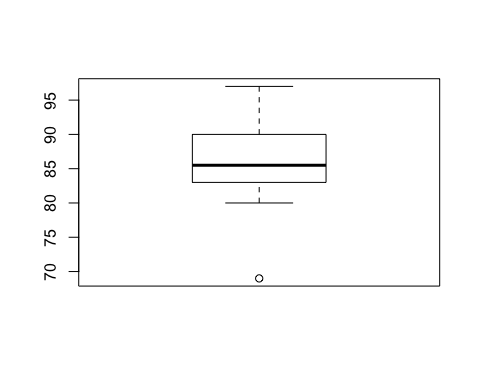
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 69.0 83.0 85.5 86.3 90.0 97.0

# make a boxplot of all the ratings together  
boxplot(ratings$Score)  
  
# there's an outlier! ratings should be between 80 and 100 inclusive  
# remove outlier  
ratings <- subset(ratings, Score >= 80)  
  
# or alternatively  
ratings <- subset(ratings, Score > 79)  
  
# install tidyverse, a collection of incredibly useful R packages  
# commenting this out for now, but do run this if you don't already have it installed  
# install.packages("tidyverse")  
  
# load tidyverse  
library(tidyverse)

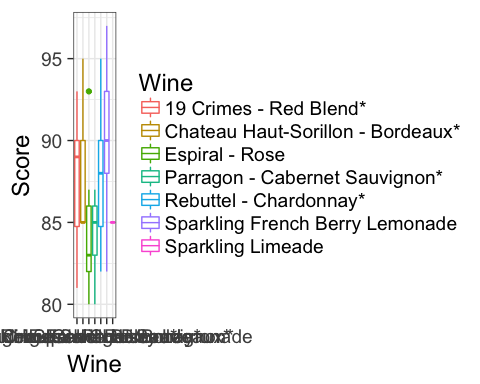
## Loading tidyverse: ggplot2  
## Loading tidyverse: tibble  
## Loading tidyverse: tidyr  
## Loading tidyverse: purrr  
## Loading tidyverse: dplyr

## Conflicts with tidy packages ----------------------------------------------

## filter(): dplyr, stats  
## lag(): dplyr, stats

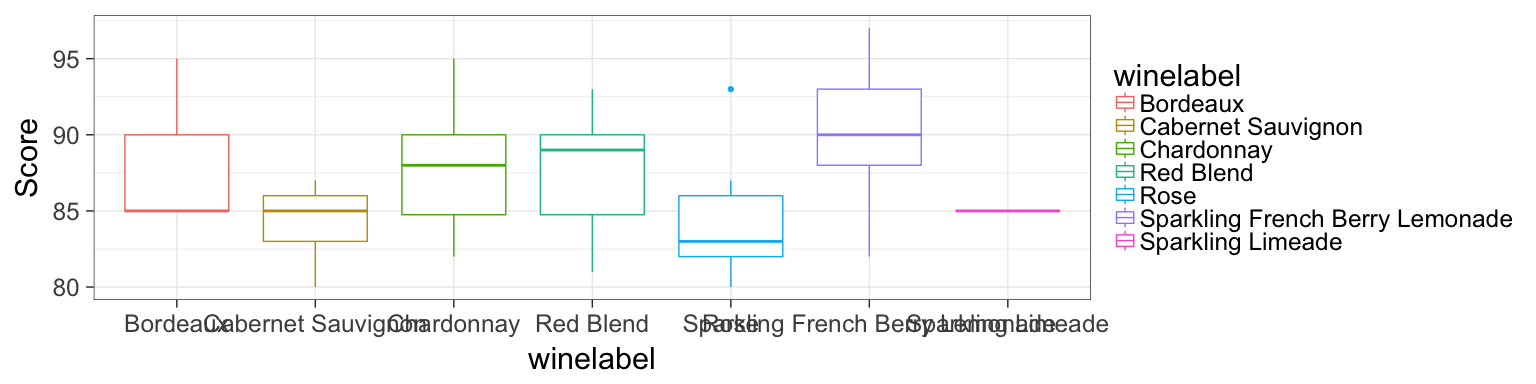


# or alternatively  
require(tidyverse)  
  
# take a look at mean rating by wine type using dplyr and save it to a dataset called ratings.means  
# look! the arrow can go both ways  
ratings %>% group\_by(Wine) %>% summarise(meanRating=mean(Score)) -> ratings.means  
  
# make a boxplot using ggplot2 with all of the data grouped by wine type  
# note that 'ratings, aes(x=Wine, y=Score, color=Wine)' could alternatively be put in geom\_boxplot  
# right now they're treated as global variables when put in the ggplot() parentheses  
# theme\_bw() makes the background white and the number sets the text size  
ggplot(ratings, aes(x=Wine, y=Score, color=Wine)) + geom\_boxplot() + theme\_bw(18)

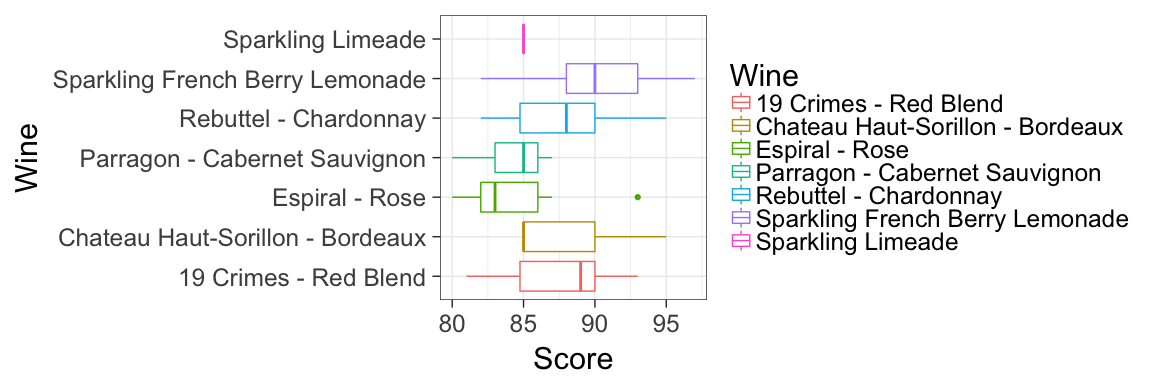


# need to clean up the x-axis so the labels are non-overlapping:  
# clean up wine label text:  
ratings$Wine <- gsub('[\*]', '', ratings$Wine)  
  
# make text shorter for display  
ratings$winelabel <- gsub('.\* - ', '', ratings$Wine)

# re-create graph  
ggplot(ratings, aes(x=winelabel, y=Score, color=winelabel)) + geom\_boxplot() + theme\_bw(23)



# still overlapping, so try a different approach and flip the boxplot  
ggplot(ratings, aes(x=Wine, y=Score, color=Wine)) + geom\_boxplot() + theme\_bw(23) + coord\_flip()



# save the output to a pdf  
pdf("/Users/Eleanor/Desktop/wineRatings.pdf", height=7, width=12)  
ggplot(ratings, aes(x=Wine, y=Score, color=Wine)) + geom\_boxplot() + theme\_bw(23) + coord\_flip()  
dev.off()

## quartz\_off\_screen   
## 2