Enterprise DNA Data Analysis Workout 005

Brian Julius

4/29/23

# Setup Chunk

#| label: setup #| include: false

library(tidyverse)
library(ggpubr)
library(ggthemes)

tips <- read\_csv("https://raw.githubusercontent.com/guipsamora/pandas\_exercises/master/07\_Visualization/Tips/tips.csv", col\_types = cols(...1 = col\_skip()))

tips$day <- factor(tips$day, levels = c("Thur", "Fri", "Sat", "Sun"))
tips$time <- factor(tips$time, levels = c("Lunch", "Dinner"))
tips$size <- factor(tips$size, levels = c( "1", "2", "3", "4", "5", "6"))

##Structure of “tips” file to be used in analysis:

# Note: Q1 incorporated in file import step

#| label: data overview
#| echo: false
knitr::kable(head(tips, 5))

| total\_bill | tip | sex | smoker | day | time | size |
| --- | --- | --- | --- | --- | --- | --- |
| 16.99 | 1.01 | Female | No | Sun | Dinner | 2 |
| 10.34 | 1.66 | Male | No | Sun | Dinner | 3 |
| 21.01 | 3.50 | Male | No | Sun | Dinner | 3 |
| 23.68 | 3.31 | Male | No | Sun | Dinner | 2 |
| 24.59 | 3.61 | Female | No | Sun | Dinner | 4 |

##

## Q2: Plot the total\_bill column histogram

gghistogram(tips,
 x="total\_bill",
 fill = "lightblue",
 bins = 15,
 rug = TRUE,
 add = "mean",
 xlab = "Total Bill",
 ylab = "Count",
 main = "Distribution of Total Bill"
 )



## Q3: Create a scatterplot presenting the relationship between total\_bill and tip

ggscatter(tips,
 x="total\_bill",
 y="tip",
 xlab = "Total Bill",
 ylab = "Tip",
 conf.int = TRUE,
 add = "reg.line",
 add.params = list(linetype = "solid", color = "red"),
 main = "Relationship Between Total Bill and Tip"
 )



## Q4: Create one image with the relationship of total\_bill, tip and size

ggscatter(data = tips,
 x = "total\_bill",
 y = "tip",
 xlab = "Total Bill",
 palette = rev( c("#d73027","#fc8d59","#fee090","#e0f3f8","#91bfdb","#4575b4")),
 ylab = "Tip",
 color = "size",
 size = "size",
 main = "Relationship of Total Bill, Tip and Size"
 ) + guides(size = guide\_none()) + labs(color = "Size") + theme(legend.position = "bottom")



## Q5: Present the relationship between days and total\_bill value, differ the dots by sex

ggstripchart(data = tips,
 palette = c("red", "blue"),
 color = "sex",
 x = "day",
 y = "total\_bill",
 xlab = "",
 ylab = "Total Bill",
 main = "Distribution of Total Bill by Day"
 ) + labs(color = "Gender") + theme(legend.position = "bottom")



## Q6: # Create a box plot presenting the total\_bill per day differentiated by time

ggboxplot(data = tips,
 x = "day",
 y = "total\_bill",
 fill = "time",
 xlab = "",
 ylab = "",
 main = "Distribution of Total Bill by Day and Time"
 ) + labs(fill = "Meal") + theme(legend.position = "bottom")



## Q7: Create two histograms (side by side) of the tip value based for Lunch and Dinner

gghistogram(tips,
 x="total\_bill",
 fill = "time",
 bins = 15,
 rug = TRUE,
 main = "Distribution of Total Bill by Time",
 facet.by = "time",
 xlab = "",
 ylab = ""
 ) + labs(fill = "Meal") + theme(legend.position = "none")



# Q8: Create two scatterplots (Male and Female) presenting the total\_bill value and tip relationship, differing by smoker/non-smoker

ggscatter(tips,
 x="total\_bill",
 y="tip",
 conf.int = TRUE,
 palette = c("cyan", "navyblue"),
 color = "smoker",
 facet.by = "sex",
 add = "reg.line",
 xlab = "Total Bill",
 ylab = "Tip",
 add.params = list(linetype = "solid", color = "red"),
 main = "Tip by Total Bill by Gender and Smoker"
 ) + labs(color = "Smoker") + theme(legend.position = "bottom")

