

# Assignment 6

## Instructions

- Please produce your assignment as a pdf (knit to pdf). See instructions in announcements if you have not downloaded a LaTeX distributor. If you are still having issues, you may knit to HTML and use your browser to produce a pdf file, as it is detailed in the M 01 09 video of module 1.
- You will have to produce this (and future) qmd file.
- You do not need to copy the statements.
- Submit your qmd file (it will not be graded but we want it for reference purposes).
- Show all the code (use `echo = TRUE` as option in R chunks) as well as the results.
- Exercises 1 and 3 are worth 25 points each while Exercise 2 is worth 50 points.
- For Exercise 1 and 2, please use the data set `consumer_complaints.csv`, which is given with this pdf in Assignment 6, and use `ggplot` only.
- We have provided `theme1` and `cbPalette`; make sure your code uses them in Exercise 1.

```
library(knitr)
library(tidyverse)
library(forcats)
library(gridExtra)
library(RSQLite)
library(jsonlite)
library(gtable)
library(grid)
library(latex2exp)
library(gridBase)
library(nnet)
library(magrittr)
library(ggplot2)
```

```
cbPalette <- c("#999999", "#E69F00", "#56B4E9", "#009E73",
               "#F0E442", "#0072B2", "#D55E00", "#CC79A7")
```

```
theme1 <- theme_bw() +
  theme(axis.text = element_text(size = 8, colour = "#6b3447"),
        axis.title = element_text(size = 10, colour = "#2f2f63"),
        legend.title = element_text(size = 8, colour = "#2f2f63"),
        legend.text = element_text(size = 8, colour = "#6b3447"),
        title = element_text(size = 12, colour = "#2f2f63"),
        axis.ticks = element_line(colour = "#6b3447"),
        plot.caption = element_text(size = 8, colour = "#2f2f63"),
        plot.subtitle = element_text(size = 10, colour = "#2f2f63"))
```

### Exercise 1: ["25" points]

Use ggplot to create multiple density plots for the number of days to resolve each consumer complaint. The number of days must be calculated in R and this variable will be called `date_diff` (`date_diff` will be the x in the density plot).

1. Set `fig.align='center'`, `fig.height=9`, `fig.width=11`.
2. Calculate the `date_diff`. (Hint: use `as.date`, `as.character`, `format="%m/%d/%Y"`, subtract `date_sent_to_company` and `date_received`)
3. Use gaussian as the kernel.
4. Label the x-axis "Number of Days" and Label the y-axis "Density Estimate".
5. Label the main title "Number of Days to Resolve Consumer Complaints".
6. Use `facet_wrap` for product and set `scales="free"`.
7. Adjust the x-axis labels to 45 degrees and have a horizontal justification of 1 (Use theme and `element_text`).
8. Use `scale_x_continuous` with limits being from 0 to 120.

Note: The majority of the data will be concentrated around 0. Do not worry about this.

```
complaints_data <- read_csv("consumer_complaints.csv")

## Begin Solution

## End Solution
```

### Exercise 2: ["50" points]

Part a: ["30" points]

1. Use ggplot to create a bar graph of product and `fill=submitted_via`.

2. Use `theme1` and set `fig.align='center',fig.height=15, fig.width=11`.
3. Use `facet_wrap` for the Year with 3 columns set to a font size 15 (Hint: Use `strip.text.x` and `element_text`).
4. Adjust the x-axis labels to 90 degrees and have a horizontal justification of 1 and set to a font size 13. Set the y-axis labels to a font size 13. Set the title, x-axis title, and y-axis title to a font size 17 (Hint: Use `theme`, `element_text`, `angle`, `hjust`, `size`).
5. Label the x-axis “Product” and the y-axis “Number of Complaints”.
6. Label the main title “Number of Complaints by Product, Submission Method, and Year”.
7. Use `scale_fill_manual` with `values=cbPalette`.

```
complaints_data <- read_csv("consumer_complaints.csv")

## Begin Solution

## End Solution
```

Part b: [“20” points]

1. Subset the data to only include the states of CA, FL, MO, and TX.
2. Use `ggplot` to create a bar graph for the number of complaints based on the submission method for the states of CA, FL, MO, and TX, where `x=submitted_via` and `fill=submitted_via` are inside `ggplot`.
3. Use `theme1` and set `fig.align='center',fig.height=9, fig.width=11`.
4. Use `facet_wrap` for state, set scales to be free in the y-axis with 4 columns, and set to a font size 15 (Hint: Use `strip.text.x` and `element_text`).
5. Adjust the x-axis labels to 35 degrees and have a horizontal justification of 1 and set to a font size 13. Set the y-axis labels to a font size 13. Set the title, x-axis title, and y-axis title to a font size 17 (Hint: Use `theme`, `element_text`, `angle`, `hjust`, `size`).
6. Label the x-axis “Submission Method” and the y-axis “Number of Complaints”.
7. Label the main title “Number of Complaints by Submission Method and State”.
8. Use `scale_fill_manual` with `values=cbPalette` and use `guides(fill="none")`.

```
## Begin Solution

## End Solution
```

### Exercise 3: ["25" points]

Using the mtcars built-in data set, plot the densities of the mpg of the different kinds of cyl on the same plot, with appropriate labeling. Add the mean of the mpg of each cylinder as a vertical dashed line. You may find it useful to look into ggplot, geom\_density, geom\_vline, and labs for this.

Additionally, please allow your code to be able to run regardless of the amount of different kinds of cylinders there are (In this case there are 3, but if given an update to the dataset to include something like 10 cylinders, it should run all the same). In other words, please avoid hardcoding if possible. You do not need to create a function for this.

HINT: The cyl variable in the dataset is a numeric type. However, for ggplot to recognize to create different densities on the same plot, the type of cyl will have to be changed.

```
## Begin Solution
```

```
## End Solution
```