STAT 201: Midterm 1 Practice 1

Possible solutions

Exercise 1

```
# note, I updated the .qmd template to ask for this.
# If you are working out of an older version, you'll filter
# out NA prices in later questions!
wine_ratings <- wine_ratings |>
  filter(!is.na(price))

# wine_ratings |>
#   count(country) |>
#   arrange(-n) |>
#   slice(1:3)

wine_ratings_top <- wine_ratings |>
  filter(country %in% c("US", "France", "Italy"))
```

Exercise 2

```
wine_ratings_top |>
  group_by(country) |>
  summarise(mean_pts = mean(points), mean_price = mean(price),
            sd_points = sd(points), sd_price = sd(price)) |>
  kable()
```

country	mean_pts	mean_price	sd_points	sd_price
France	88.69663	47.95506	3.227371	103.75807
Italy	88.31579	44.09868	2.743858	41.70327
US	88.37030	35.64098	3.150957	25.72696

Answer:

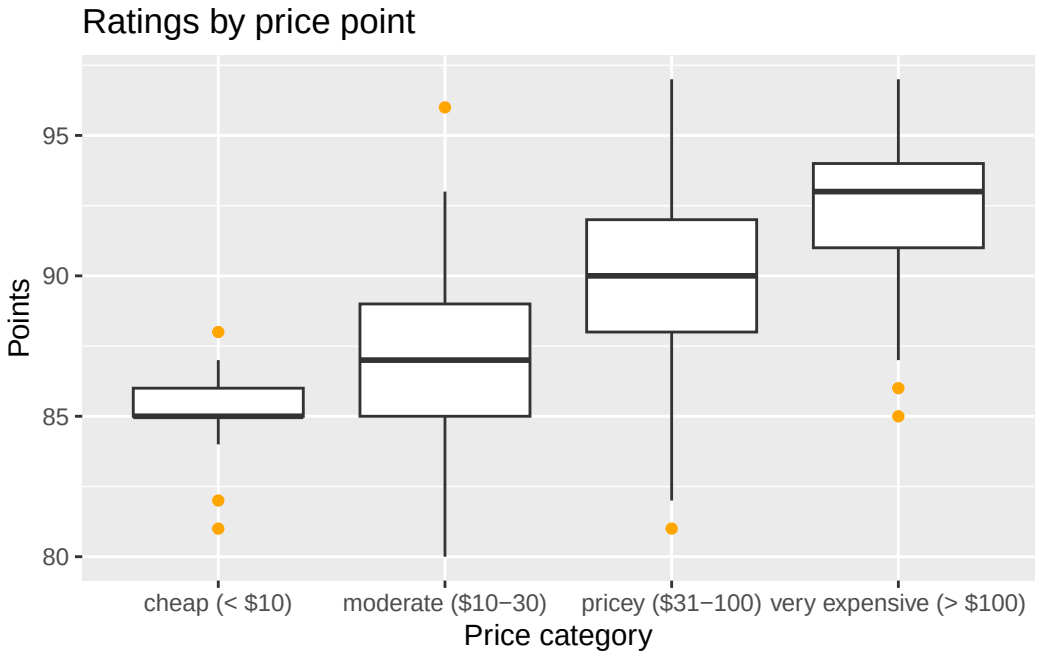
Exercise 3

```
wine_ratings_top |>
  filter(year < 2010) |>
  mutate(over90 = if_else(points > 90, T, F)) |>
  count(over90, country) |>
  group_by(country) |>
  mutate(prop = n/sum(n)) |>
  ungroup() |>
  filter(over90 == TRUE) |>
  select(country, prop) |>
  kable()
```

country	prop
France	0.3947368
Italy	0.2909091
US	0.2055556

Exercise 4

```
wine_ratings |>
  mutate(price_cat = case_when(
    price < 10 ~ "cheap (< $10)",
    price >= 10 & price <= 30 ~ "moderate ($10-30)",
    price > 30 & price <= 100 ~ "pricey ($31-100)",
    price > 100 ~ "very expensive (> $100)"
  )) |>
  ggplot(aes(x = price_cat, y = points)) +
  geom_boxplot(outlier.color = "orange") +
  labs(x = "Price category", y = "Points", title = "Ratings by price point")
```



Exercise 5

```
## option 1
most_reviewed <- wine_ratings |>
  count(variety) |>
  filter(n >= 25)

wine_ratings |>
  mutate(point_price = points/price) |>
  inner_join(most_reviewed, by = "variety") |>
  group_by(variety) |>
  summarise(mean = mean(point_price)) |>
  arrange(mean) |>
  slice(1)
```

```
# A tibble: 1 x 2
  variety     mean
  <chr>      <dbl>
1 Pinot Noir  2.53
```

```r
## option 2
wine_ratings |>
  mutate(point_price = points/price) |>
  group_by(variety) |>
  summarise(mean = mean(point_price), n = n()) |>
  arrange(-n) |>
  filter(n >= 25) |>
  arrange(mean) |>
  slice(1)
```

```
# A tibble: 1 x 3
  variety     mean     n
  <chr>      <dbl> <int>
1 Pinot Noir  2.53   117
```