L04 ggplot III

Data Visualization (STAT 302)

Contents

Shay Lebovitz

Overview

The goal of this lab is to continue the process of unlocking the power of `ggplot2` through constructing and experimenting with a few basic plots.

Datasets

We'll be using data from the `blue_jays.rda`, `cows.rda`, and `cdc.txt` datasets which are already in the `/data` subdirectory in our **data__vis__labs** project.

```
# Load package(s)
library(tidyverse)

# Load datasets
load('data/blue_jays.rda')
load('data/cows.rda')
cdc <- read_delim(file = "data/cdc.txt", delim = "|") %>%
  mutate(genhlth = factor(genhlth,
                          levels = c("excellent", "very good", "good", "fair", "poor")
  ))

# Read in the cdc dataset
cdc <- read_delim(file = "data/cdc.txt", delim = "|") %>%
  mutate(genhlth = factor(genhlth,
    levels = c("excellent", "very good", "good", "fair", "poor")
  ))
```

Complete the following exercises.

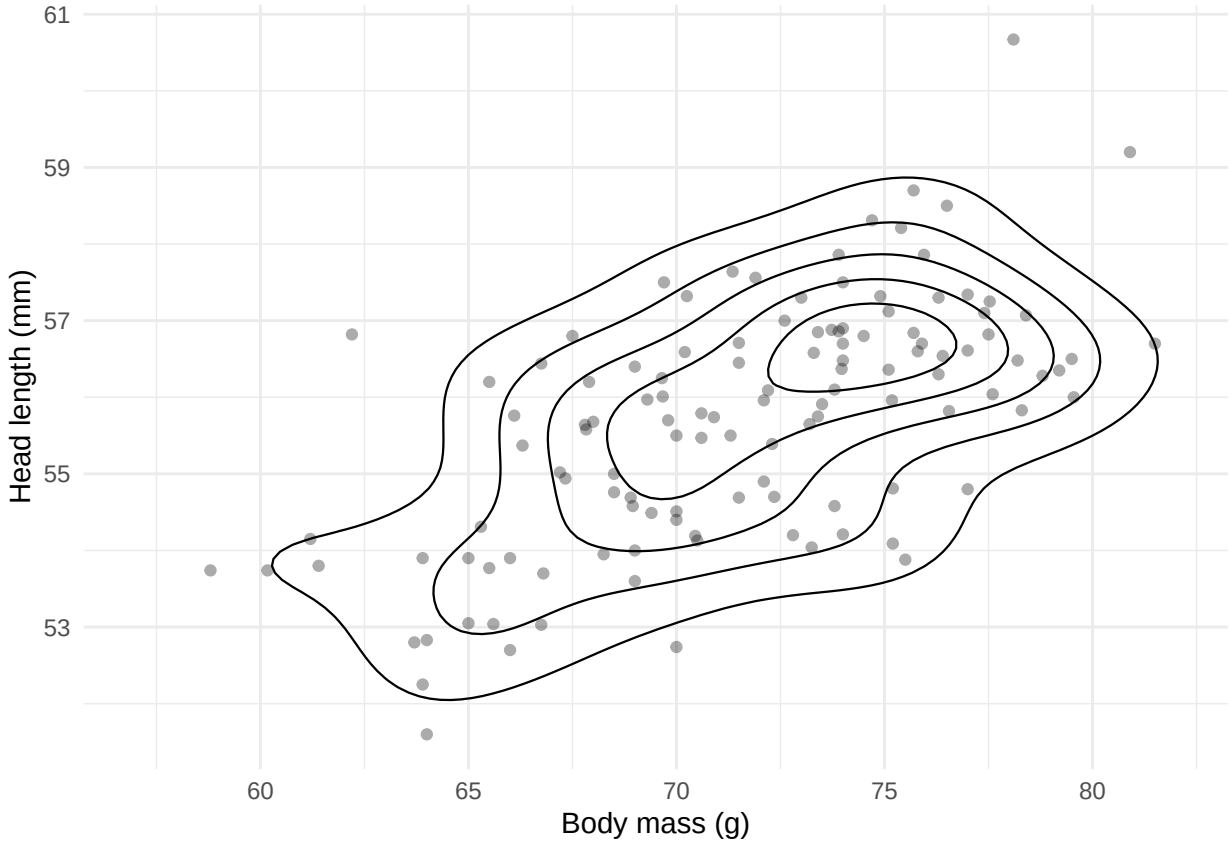
Exercise 1

Using `blue_jays.rda` dataset, recreate the following graphic as precisely as possible.

Hints:

- `geom_density_2d()` or `stat_density_2d()`
- Transparency is 1/3
- Horizontal lower limit is 57 and upper limit is 82
- Point size 1.5
- Line size is 0.4
- `binwidth` set to 0.004
- Minimal theme

```
blue_jays %>%
  ggplot(aes(x = Mass, y = Head)) +
  geom_point(alpha = 1/3, size = 1.5) +
  labs(x = 'Body mass (g)', y = 'Head length (mm)') +
  xlim(57, 82) +
  geom_density_2d(color = 'black', binwidth = 0.004, size = 0.4) +
  theme_minimal()
```



Exercise 2

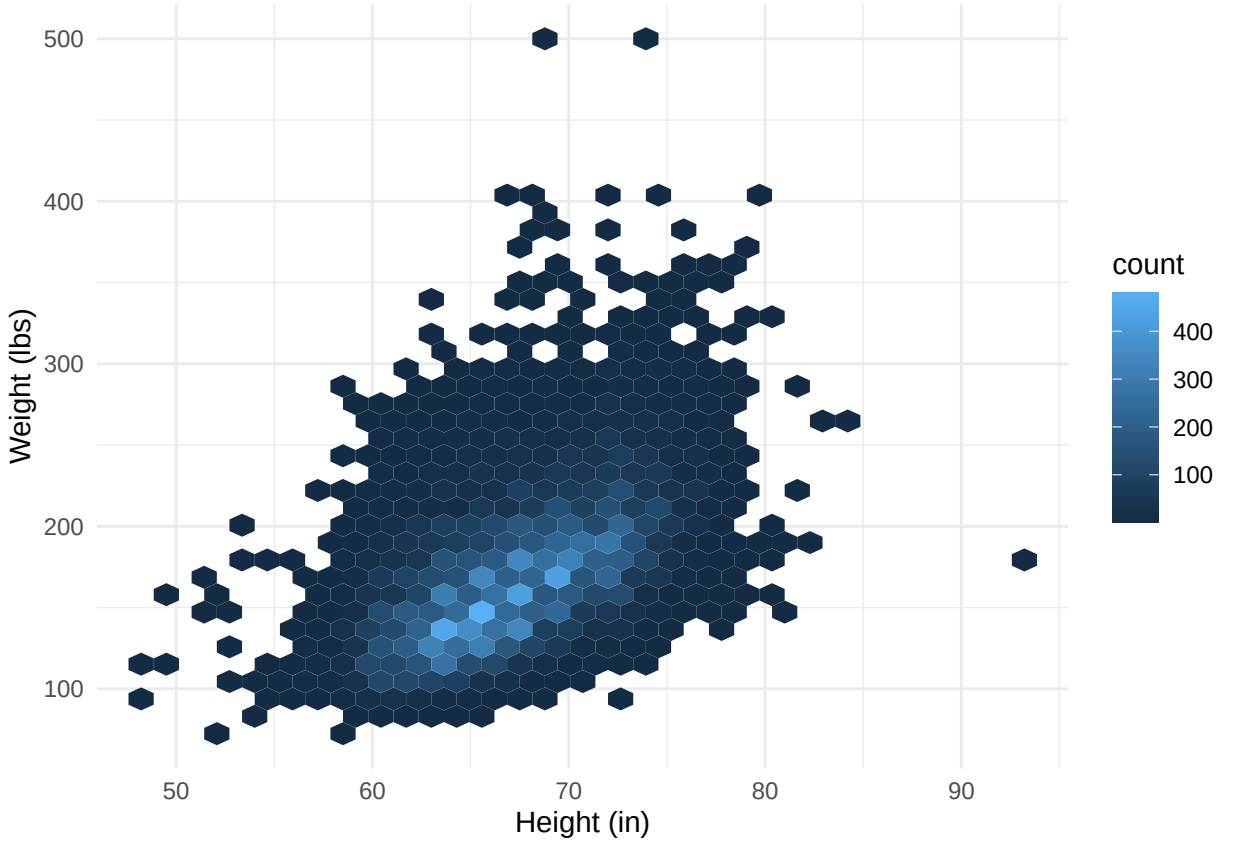
Using `cdc` dataset, recreate the following graphics as precisely as possible.

Plot 1 *Hints:*

- `bins` set to 35
- Minimal theme

```
cdc %>%
  ggplot(aes(x = height, y = weight)) +
```

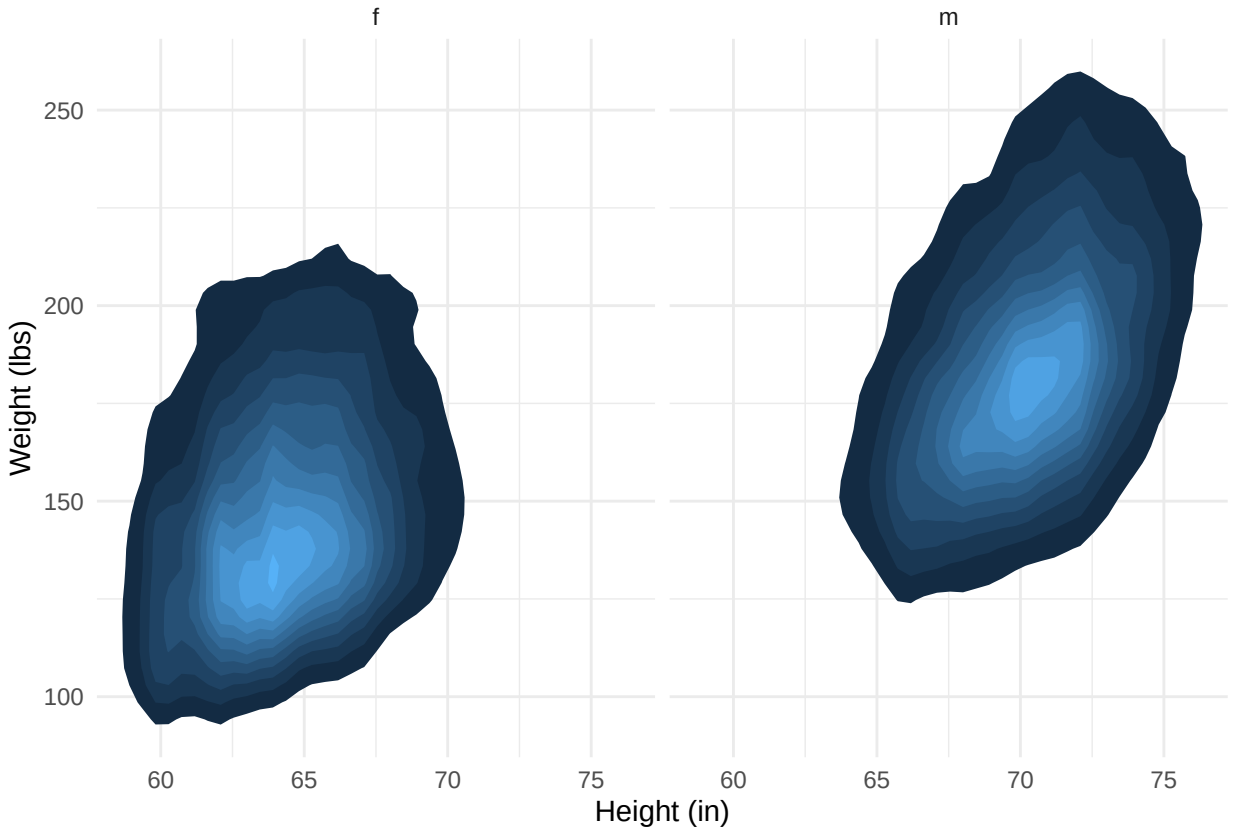
```
  geom_hex(bins = 35) +
  labs(x = 'Height (in)', y = 'Weight (lbs)') +
  theme_minimal()
```



Plot 2 *Hints:*

- `polygon`
- Minimal theme

```
cdc %>%
  ggplot(aes(x = height, y = weight)) +
  stat_density_2d(geom = 'polygon', aes(fill = ..level..)) +
  facet_wrap(~gender) +
  labs(x = 'Height (in)', y = 'Weight (lbs)') +
  theme_minimal() +
  theme(legend.position = "none")
```



Exercise 3

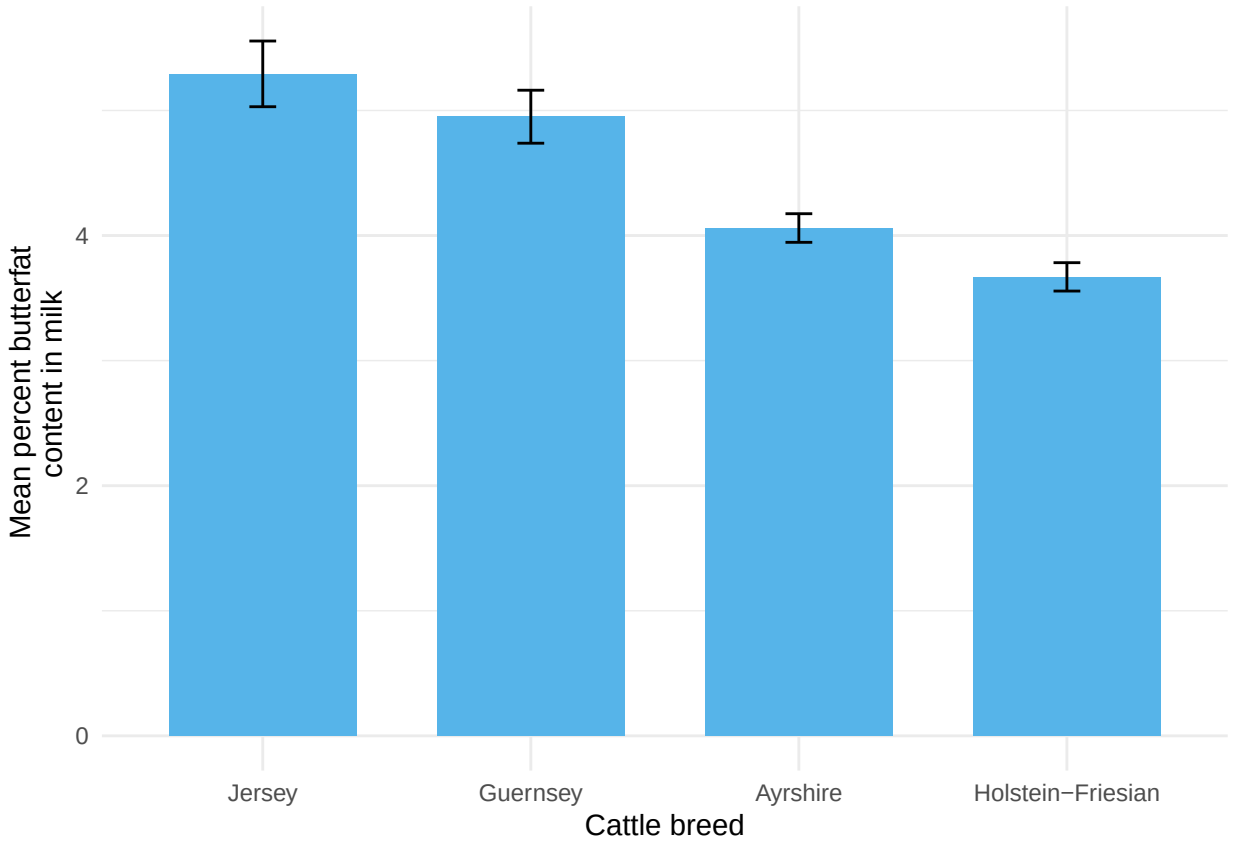
Using `cow_means` dataset derived from the `cows` dataset, recreate the following graphic as precisely as possible.

Hints:

- Hex color code `#56B4E9`
- 95% confidence intervals (1.96 or `qnorm(0.975)`)
- Some useful values: 0.1, 0.7

```
cow_means <- cows %>%
  filter(breed != "Canadian") %>%
  group_by(breed) %>%
  summarize(
    mean = mean(butterfat),
    se = sd(butterfat) / sqrt(n())
  ) %>%
  mutate(breed = fct_reorder(factor(breed), desc(mean)))
```

```
cow_means %>%
  ggplot(aes(x = breed, y = mean)) +
  geom_bar(stat = 'identity', fill = '#56B4E9', width = 0.7) +
  geom_errorbar(aes(ymin = mean - 1.96 * se, ymax = mean + 1.96 * se),
                width = 0.1) +
  labs(x = 'Cattle breed', y = 'Mean percent butterfat \n content in milk') +
  theme_minimal()
```



Exercise 4

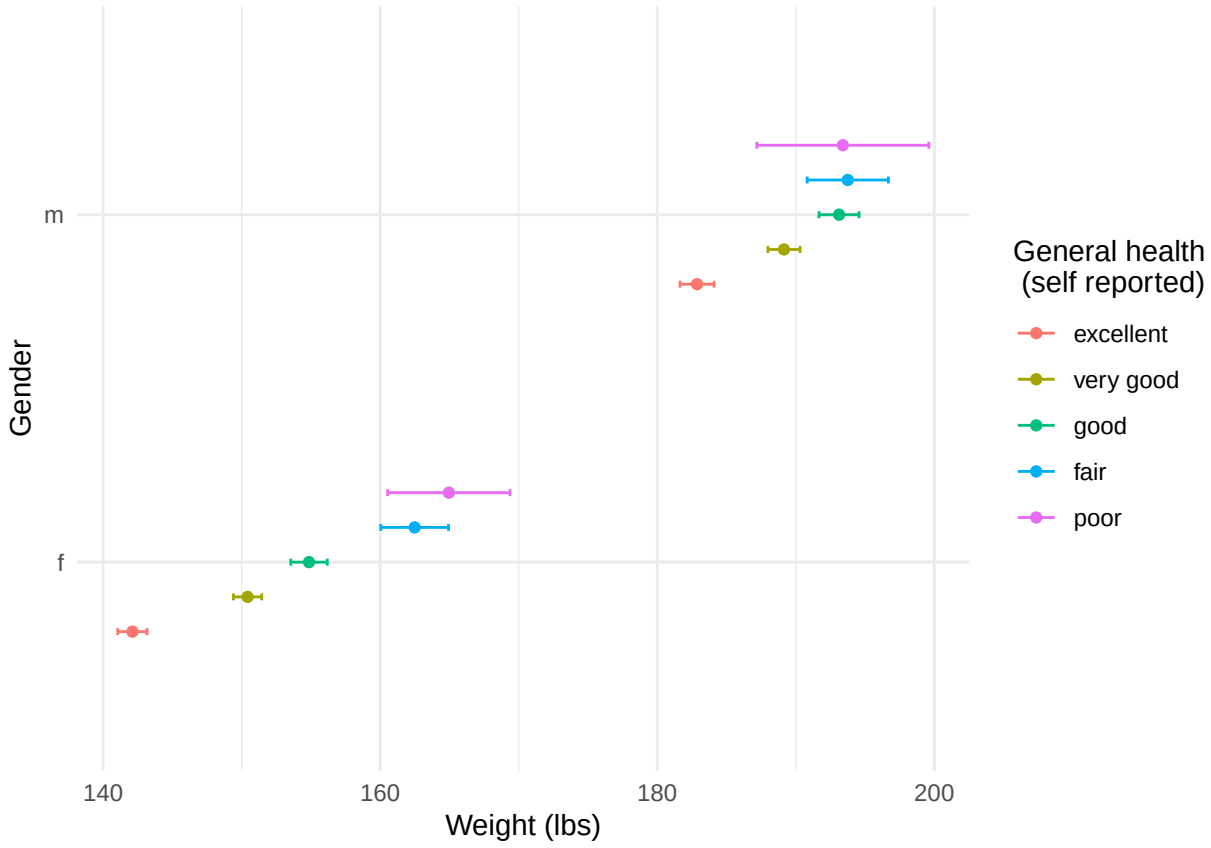
Using `cdc_weight_95ci` dataset derived from the `cdc` dataset, recreate the following graphic as precisely as possible.

Hints:

- Useful values: 0.1, 0.5

```
# 95% CI for weight for genhlth, gender groups
cdc_weight_95ci <- cdc %>%
  group_by(genhlth, gender) %>%
  summarise(
    mean_wt = mean(weight),
    se = sd(weight) / sqrt(n()),
    moe = qt(0.975, n() - 1) * se
  )
```

```
cdc_weight_95ci %>%
  ggplot(aes(x = mean_wt, y = gender, color = genhlth),) +
  geom_point(position = position_dodge(0.5)) +
  geom_errorbar(aes(xmin = mean_wt - moe, xmax = mean_wt + moe),
                width = 0.1, position = position_dodge(0.5)) +
  labs(x = 'Weight (lbs)', y = 'Gender') +
  theme_minimal() +
  labs(color = 'General health \n (self reported)')
```


Gender
m
f
140
160
180
200
Weight (lbs)
General health
(self reported)
excellent
very good
good
fair
poor