L04 ggplot III
Data Visualization (STAT 302)

Anna Wagman

Contents

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.

```r
# Load package(s)
library(tidyverse)
library(skimr)
library(hexbin)

# Load datasets
load(file = "data/blue_jays.rda")
load(file = "data/cows.rda")



# 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")
  ))
```

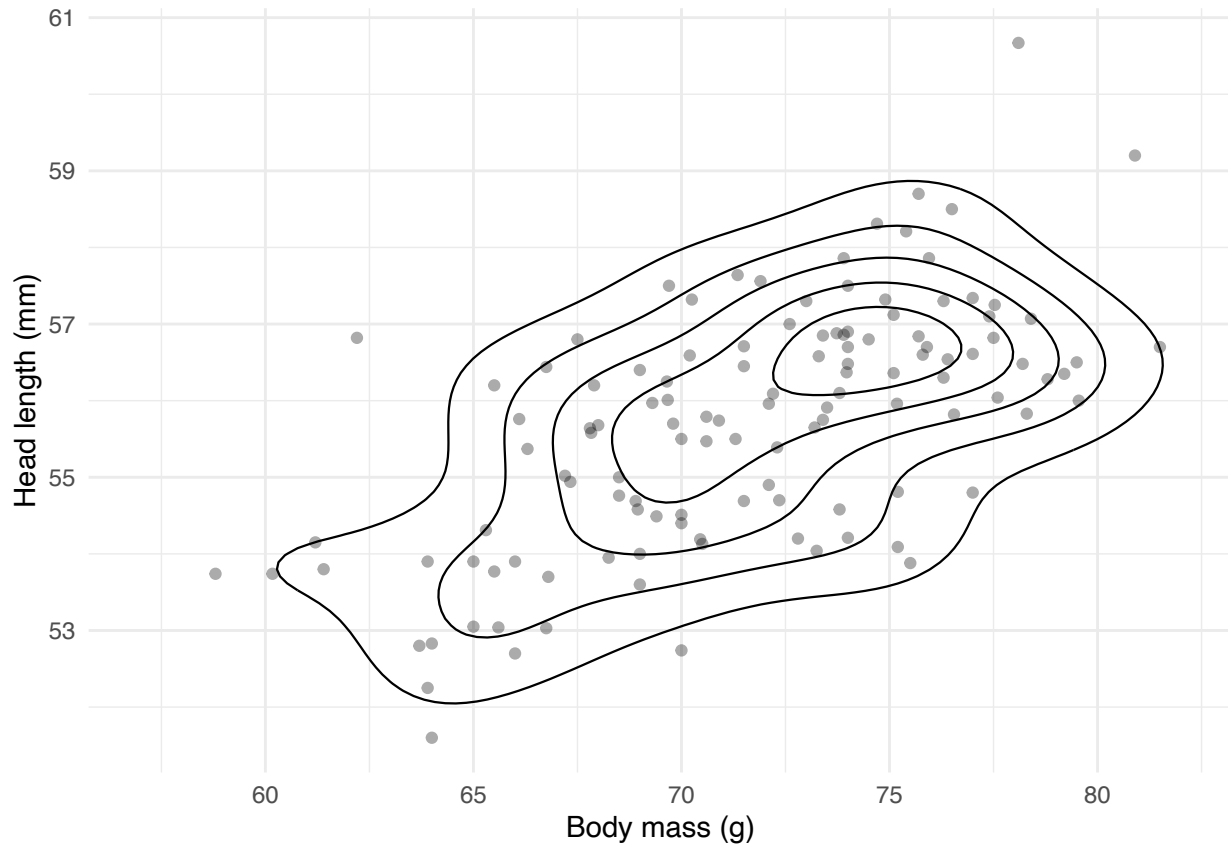Complete the following exercises.

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

```
ggplot(data = blue_jays, aes(x = Mass, y = Head)) +
  ##scatter plot points
  geom_point(alpha = 1/3, size = 1.5) +
  ##density
  geom_density_2d(
    ##set given binwidth and line size
    binwidth = 0.004, color = "black", size = 0.4
  ) +
  ##horizontal axes (x-axis) limit
  xlim(57, 82) +
  ##axes titles
  labs(
    x = "Body mass (g)",
    y = "Head length (mm)"
  ) +
  theme_minimal()
```
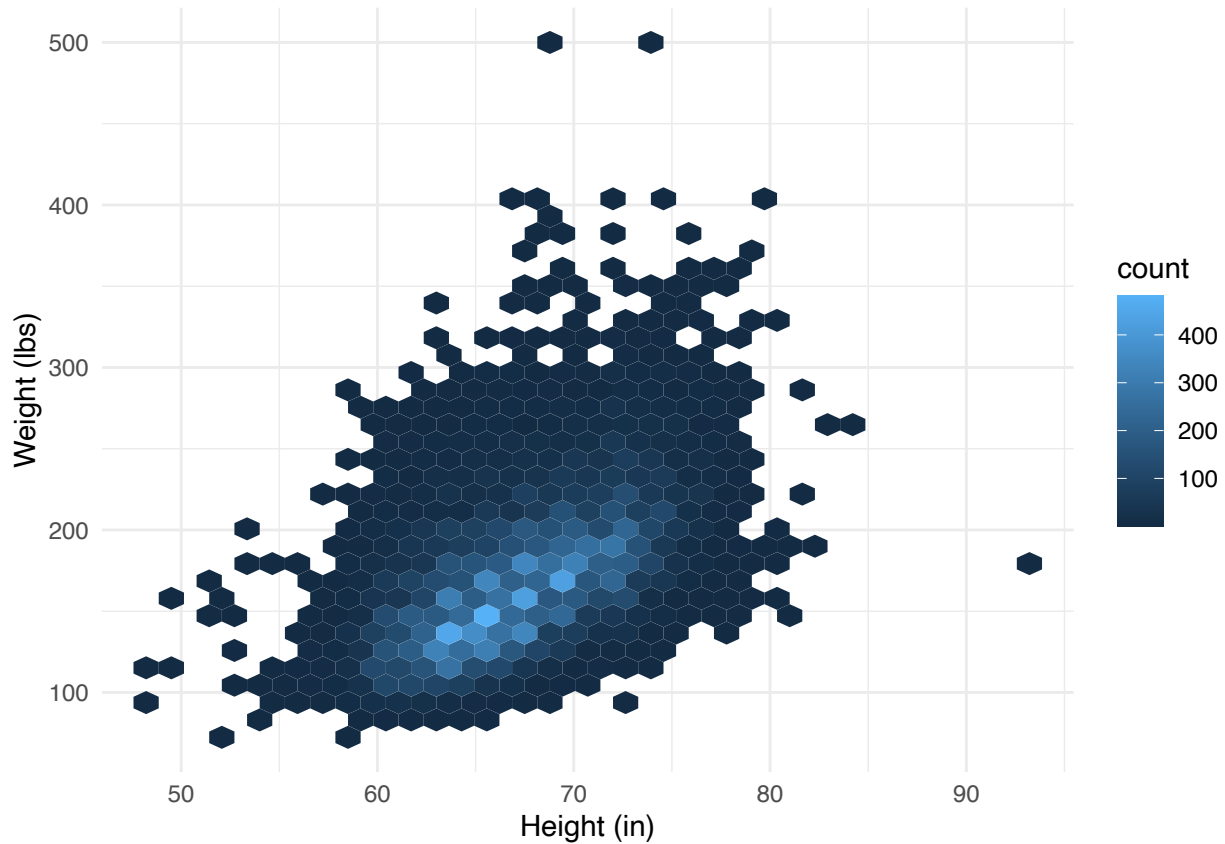

Exercise 2

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

Plot 1 *Hints:*

- `bins` set to 35

- Minimal theme

```
ggplot(
  data = cdc, aes(x = height, y = weight)) + #plot height v weight
  geom_hex(bins=35) + #hint: bins = 35
  labs(x = "Height (in)", y = "Weight (lbs)") + #label axes
  theme_minimal()
```
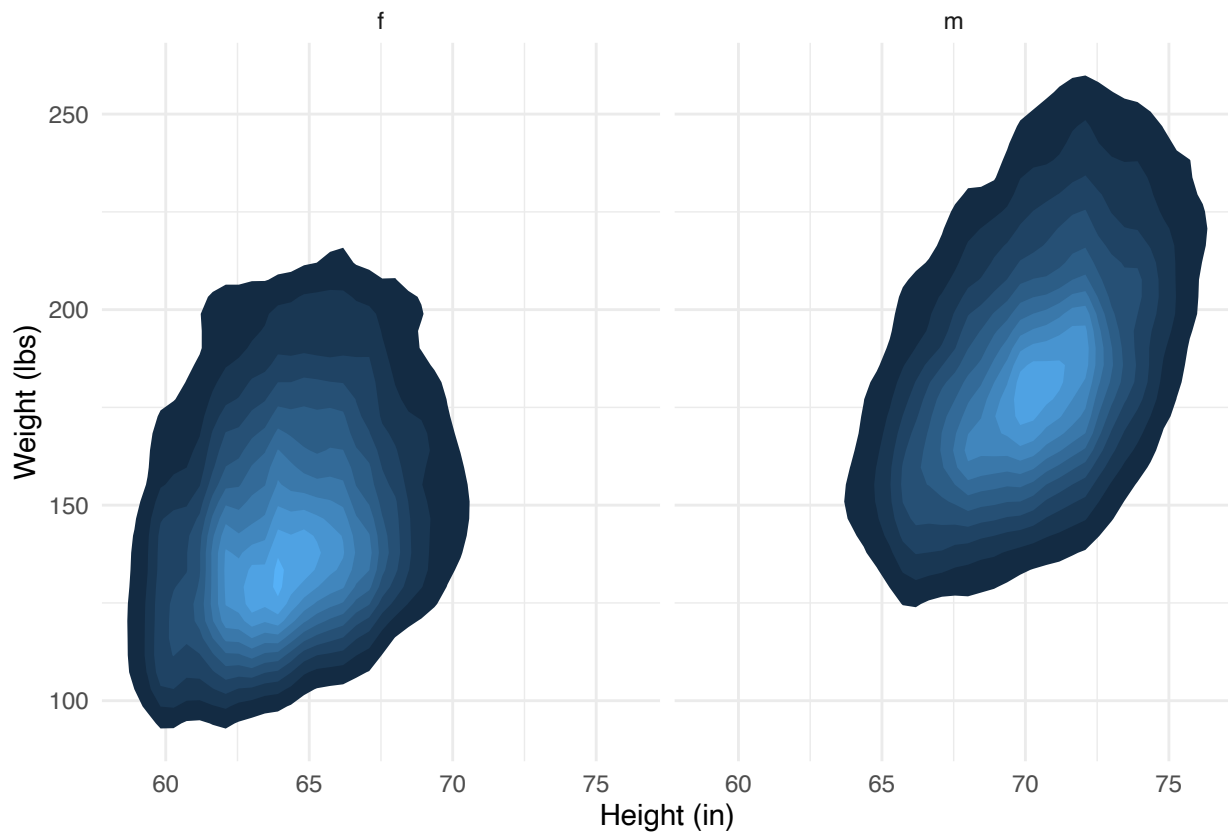

Plot 2   *Hints:*
- polygon (not a geom!)
- Minimal theme

```
ggplot(
  data = cdc, aes(x = height, y = weight)) + #plot height v weight
  labs(
    x = "Height (in)", y = "Weight (lbs)") + #axes labels

  stat_density_2d(
    aes(fill = ..level..),
    geom = "polygon",  #hint geom = polygon
    show.legend = FALSE
  ) +

  facet_wrap(~gender) + #seperate by gender

  theme_minimal()
```

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

```r
# Graphic dataset

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)))

z_star <- qnorm(0.975) #confidence interval of 95%

ggplot(
  data = cow_means,
  aes(x = breed, y = mean)) + #breed v mean

  geom_col(fill = "#56B4E9", width = 0.7) + #hex color code = #56B4E9
```
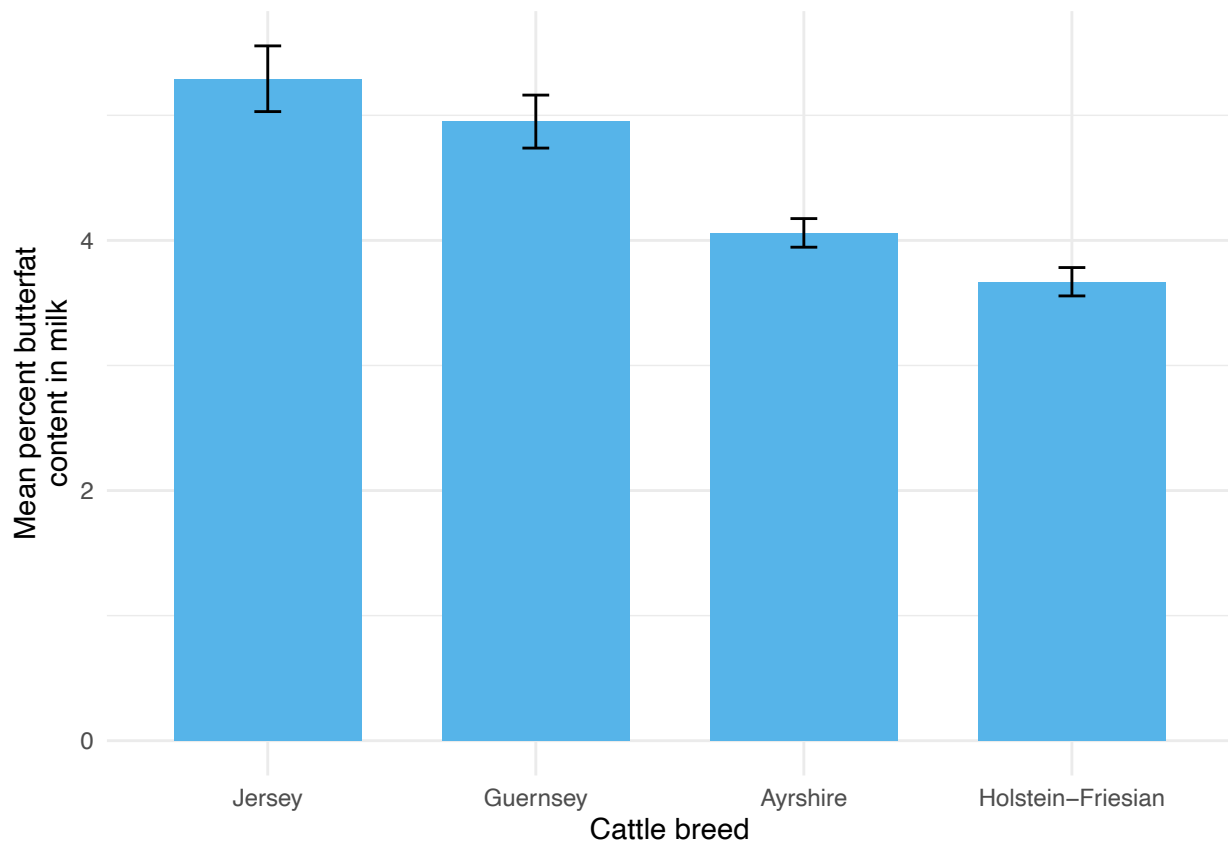
```
geom_errorbar(
  aes(ymin = mean - se*z_star,
      ymax = mean + se*z_star), width = 0.1
) +
theme_minimal() +
labs( #axes label
  x = "Cattle breed",
  y = "Mean percent butterfat\ncontent in milk"
)
```


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
  )
```

```
ggplot(cdc_weight_95ci, aes(mean_wt, gender, color = genhlth))+
  geom_point(position = position_dodge(width = 0.5)) +
  geom_errorbarh(aes(xmin = mean_wt - moe, xmax = mean_wt + moe, height = 0.1), position = position_dod
  theme_minimal() +
  labs(x = "Weight (lbs)", y = "Gender") +
  scale_color_discrete(name = "General health\n(self reported)")
```