

Lab 2 Assignment

Summer Heschong

2025-01-24

```
#Setup: load packages and data
```

```
library(here)
```

```
## here() starts at /Users/summerheschong/Documents/GitHub/stats_spring25
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
## v dplyr      1.1.4      v readr      2.1.5
```

```
## v forcats    1.0.0      v stringr    1.5.1
```

```
## v ggplot2    3.5.1      v tibble     3.2.1
```

```
## v lubridate  1.9.4      v tidyr      1.3.1
```

```
## v purrr      1.0.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(viridis)
```

```
## Loading required package: viridisLite
```

```
library(dplyr)
```

```
crab_data <- read_csv(here('Data/Raw/pie_crab.csv'))
```

```
#Part 1: Measures of location and data spread
```

```
#a. What is the mean fiddler crab carapace size?
```

```
mean(crab_data$size)
```

```
## [1] 14.65801
```

```
#The mean fiddler crab carapace size is 14.66
```

```
#b. What is the median fiddler crab carapace size?
```

```
median(crab_data$size)
```

```
## [1] 14.435
```

```
#The median fiddler crab carapace size is 14.44
```

```
#c. What is the standard deviation in fiddler crab carapace size?
```

```
sd(crab_data$size)
```

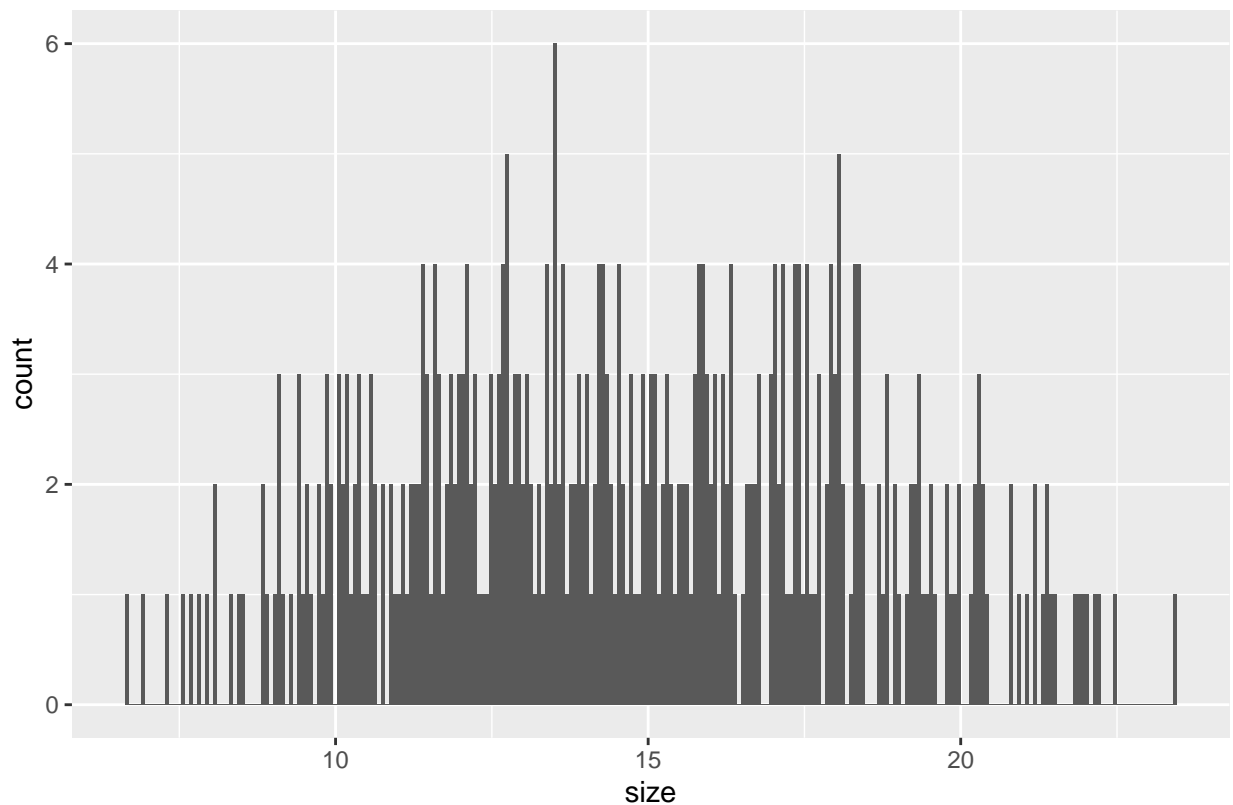
```
## [1] 3.503304
```

```
#The standard deviation in fiddler crab carapace size is 3.5
```

```
#Part 2: Creating Histograms
```

```
#a. Create histogram of fiddler crab carapace size
```

```
fig1 <- ggplot(data = crab_data, aes(x = size)) +  
  geom_histogram(bins = 261, binwidth = 0.064) +  
  labs(caption = 'Histogram of fiddler crab carapace size across sites')  
print(fig1)
```



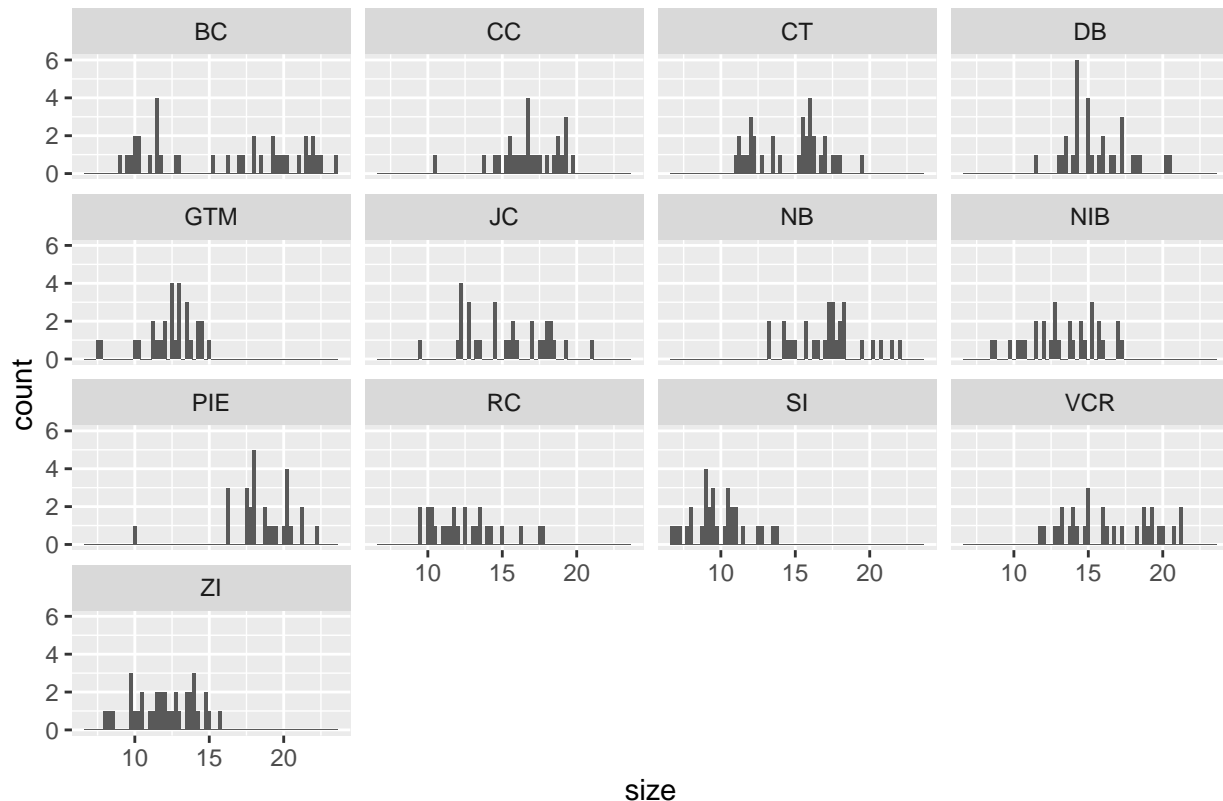
Histogram of fiddler crab carapace size across sites

```

#To find the right number of bins I used  $2n^{1/3}$  or  $2(392)^{1/3}$ ,
#and for bin width I used  $\text{range}/\text{\#of bins}$ 

#b. Create histograms of fiddler crab carapace size for each of the 13 sites
fig2 <- ggplot(data = crab_data, aes(x = size)) +
  geom_histogram(bins=20, binwidth = 0.25) +
  facet_wrap(~site) +
  labs(caption = 'Histograms of fiddler crab carapace size by site')
print(fig2)

```



Histograms of fiddler crab carapace size by site

c. How do the raw values of fiddler crab carapace size in your first figure compare to the mean and median values you calculated above? Based on your second figure, do any sites display strong differences in their data distributions?

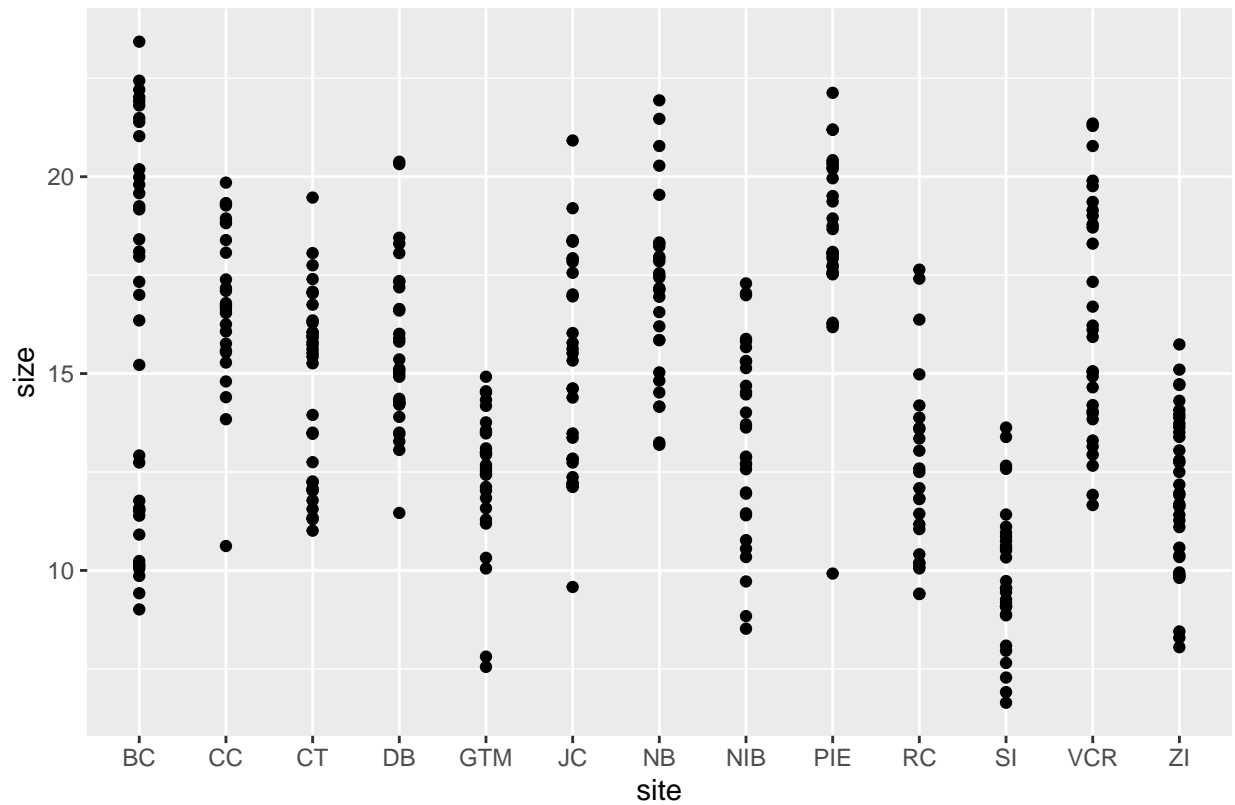
Answer: I think the raw values fit pretty well with the mean and median values I calculated above. In the first figure the values from around 12 - 16 all look pretty similar and there's a spike around 14. In the second figure all of the sites have pretty different data distributions. For example, at BC the data is pretty spread out across sizes, and at PIE the data is all to the left of size 15.

#Part 3: Data Visualization - the Golden Rules

```

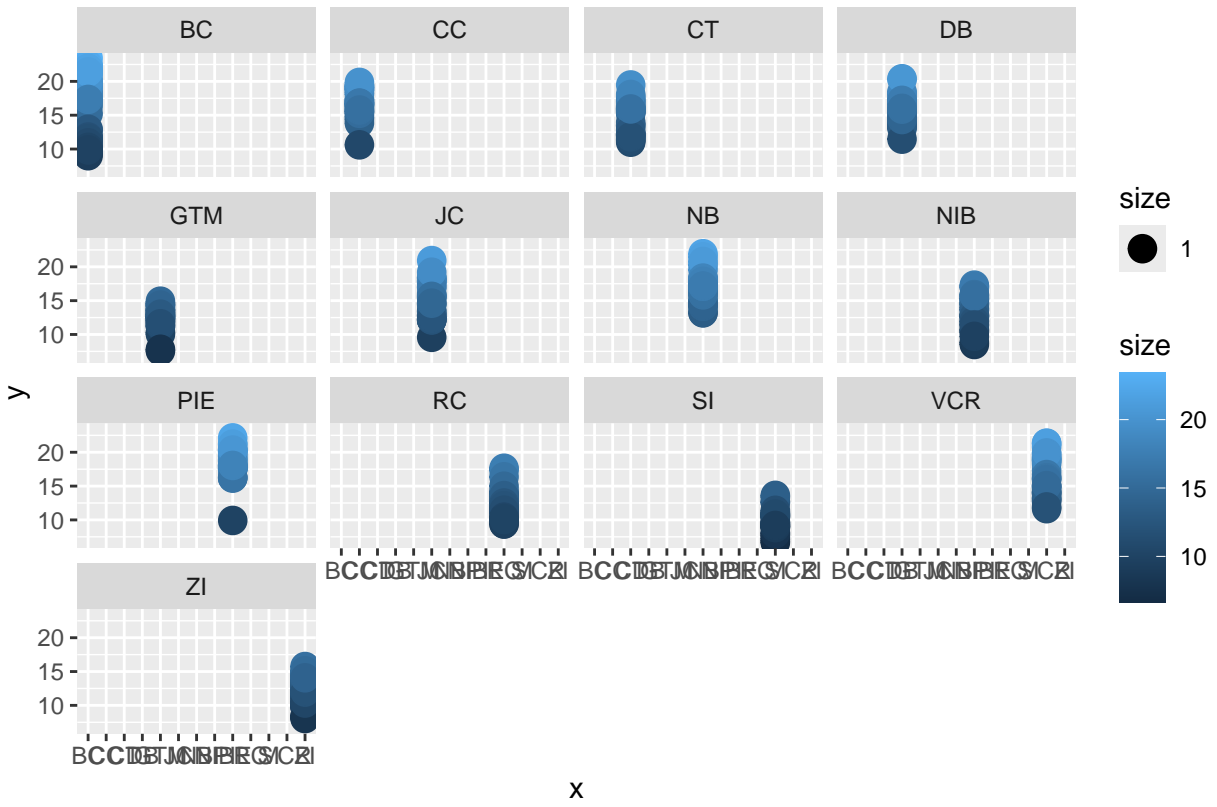
#a. Make a plot of your choice that follows the Golden Rules
fig3 <- ggplot(data = crab_data, aes(x = site, y = size)) +
  geom_point() +
  labs(caption = "Scatterplot of fiddler crab carapace size by site")
print(fig3)

```



Scatterplot of fiddler crab carapace size by site

```
#b. Create worst version of fig3
fig4 <- ggplot(
  data = crab_data,
  aes(x = site, y = size,
      color = size, size = 1 )) +
  geom_point() +
  facet_wrap(~site) +
  labs(caption = 'Carapace and island', x = 'x', y = 'y')
print(fig4)
```



Carapace and island