

# Mini-homework 6: Analyzing data distributions

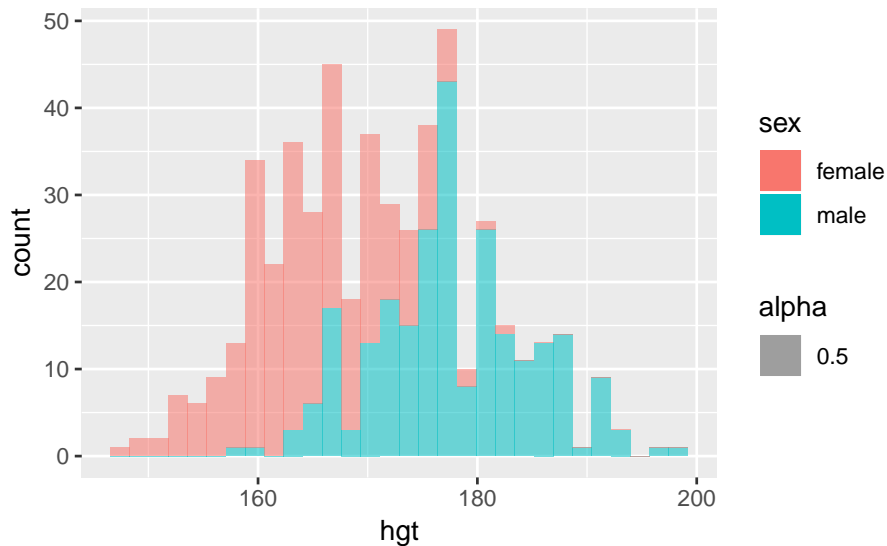
*Riley Payung*

*2019-10-21*

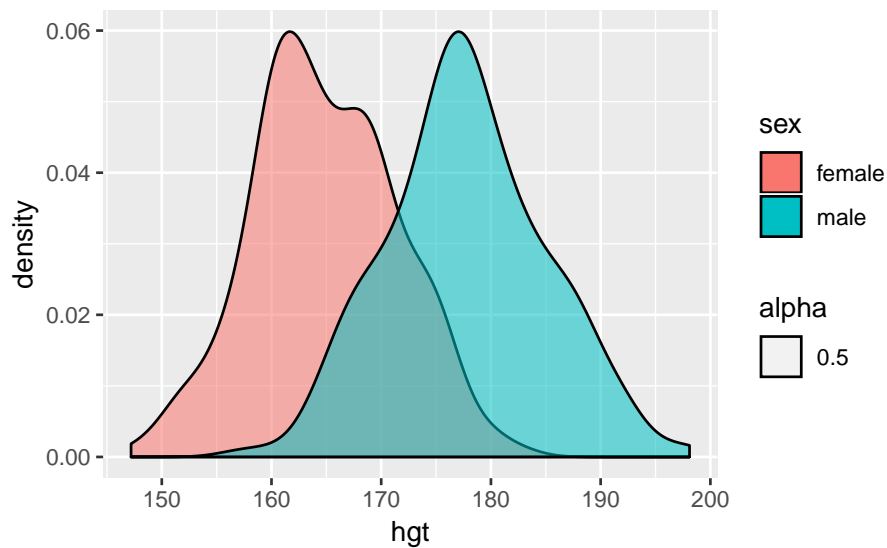
## Exercise 1

```
ggplot(body_dims) +  
  geom_histogram(mapping=aes(x = hgt, fill=sex, alpha = 0.5))
```

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



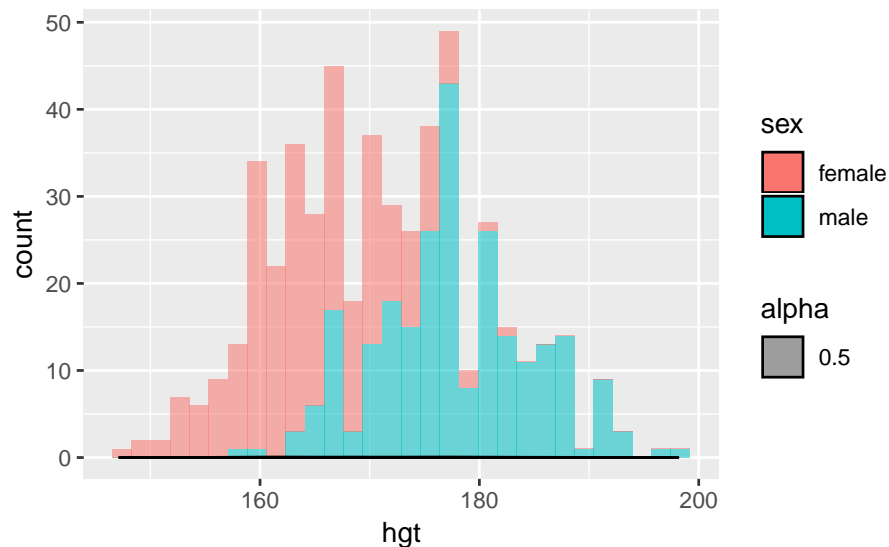
```
ggplot(body_dims) +  
  geom_density(mapping=aes(x=hgt, fill=sex, alpha = 0.5))
```



## Exercise 2

```
body_dims %>%  
  ggplot(mapping = aes(x = hgt, fill = sex, alpha = 0.5)) +  
  geom_histogram() +  
  geom_density()
```

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

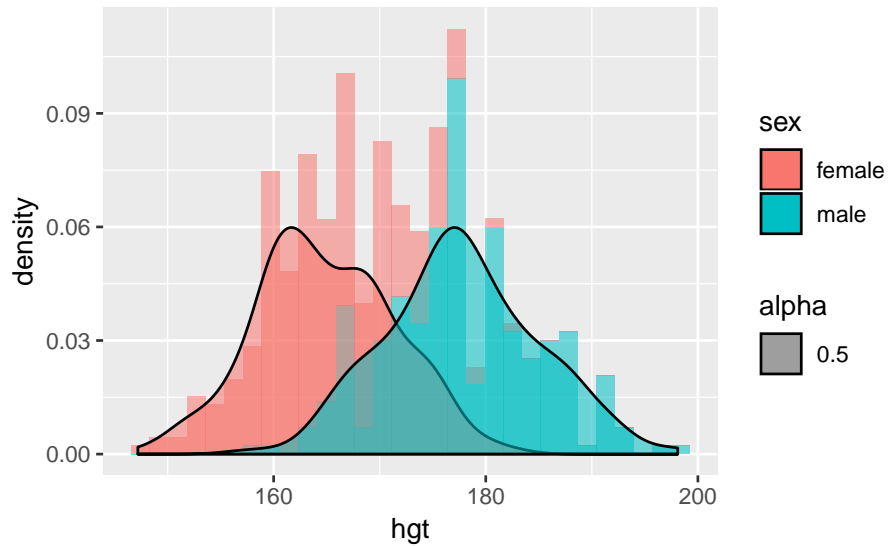


The density plot doesn't actually work with the histogram plot, therefore there is a problem. It seems like it has to do with the y-axis, since the density is relative frequency and the count is just a finite number.

## Exercise 3

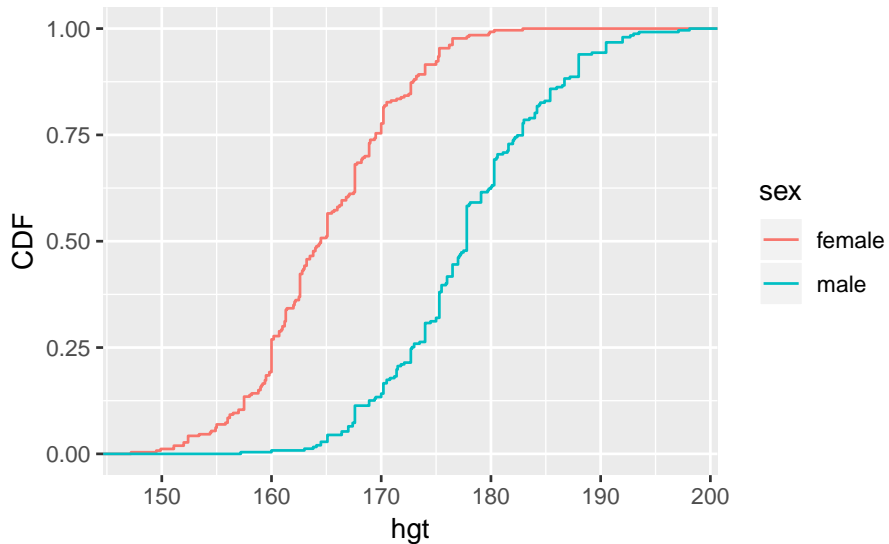
```
body_dims %>%  
  ggplot(mapping = aes(x = hgt, y = ..density.., fill = sex, alpha = 0.5)) +  
  geom_histogram() +  
  geom_density()
```

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



#### Exercise 4

```
ggplot(body_dims) +
  geom_step(
    mapping = aes(x = hgt, color=sex),
    stat = "ecdf"
  ) +
  labs(y = "CDF")
```



#### Exercise 5

It looks similar, the amount of data in the women's cdf reaches the total women in the dataset earlier than the men's. The mens also starts later. It would mostly overlap if we were to shift the men's graph to the left, with some minor deviations.

## Exercise 6

```
body_dims_summary <- body_dims %>%  
  group_by(sex) %>%  
  summarize(  
    mean = mean(hgt),  
    median = median(hgt),  
    min = min(hgt),  
    max = max(hgt),  
    sd = sd(hgt),  
    iqr = IQR(hgt)  
  )
```

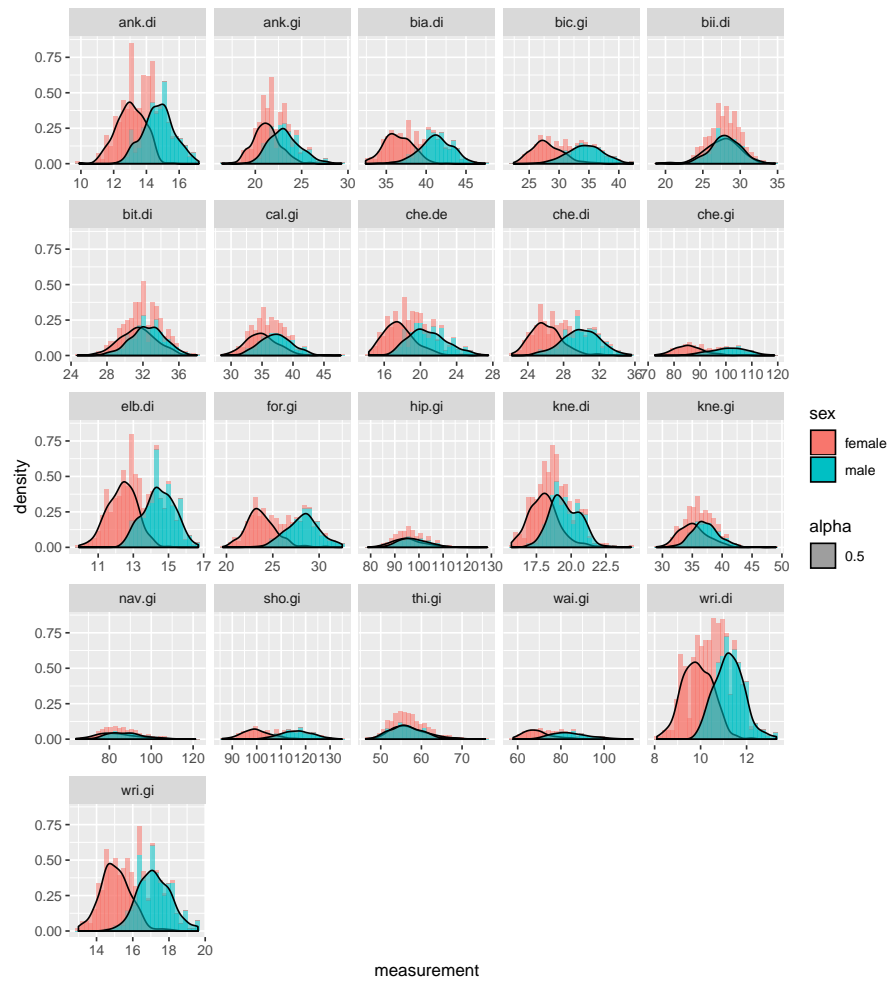
## Exercise 7

```
body_dims_long <- body_dims %>%  
  gather(  
    bia.di:wri.gi,  
    key = body_part,  
    value = measurement  
  )
```

## Exercise 8

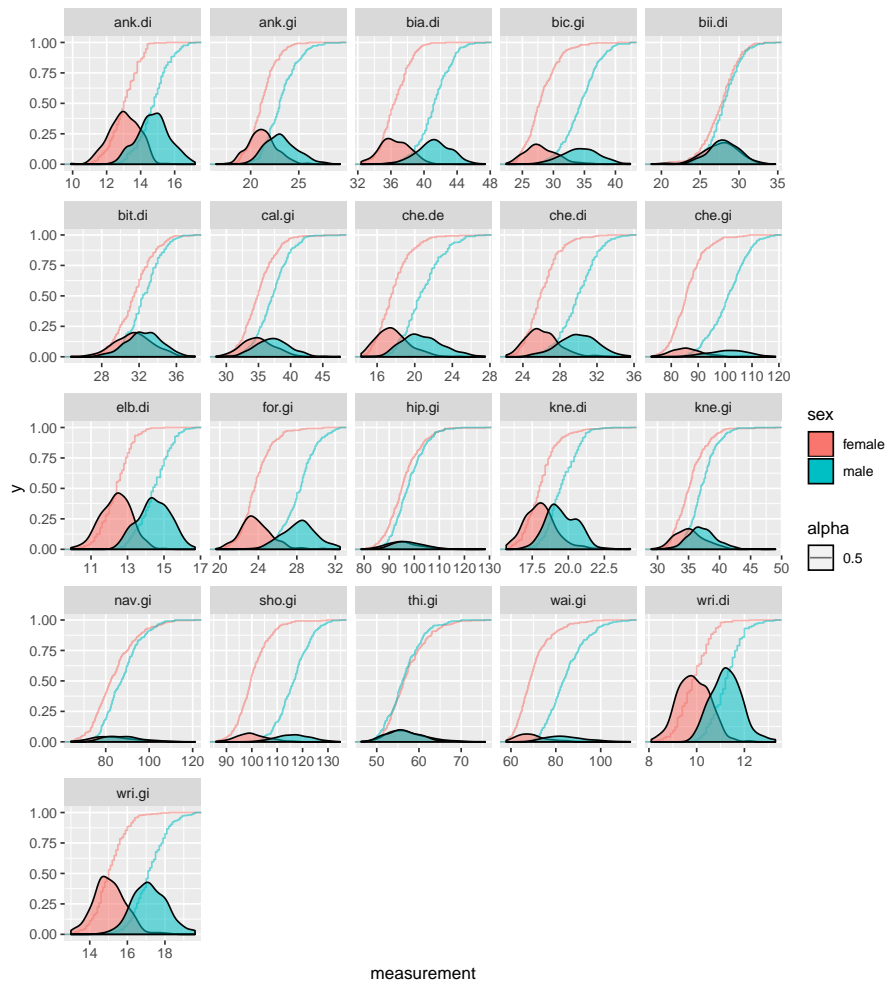
```
body_dims_long %>%  
  ggplot(mapping = aes(x = measurement, y = ..density.., fill = sex, alpha = 0.5)) +  
  geom_histogram() +  
  geom_density() +  
  facet_wrap(~ body_part, scales = "free_x")
```

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



## Exercise 9

```
ggplot(body_dims_long) +
  geom_step(mapping=aes(x = measurement, color = sex, alpha = 0.5), stat='ecdf') +
  geom_density(mapping = aes(x = measurement, y = ..density.., fill = sex, alpha = 0.5)) +
  facet_wrap(~ body_part, scales = "free_x")
```



The following are nearly indistinguishable between men and women:

1. biiliac diameter
2. hip girth
3. thigh girth

Some that are close, but are distinguishable between men and women:

1. bitrochanteric diameter
2. knee girth
3. navel (abdominal) girth

## Exercise 10

```
body_dims_summary <- body_dims_long %>%
  group_by(body_part, sex) %>%
  summarize(
    mean = mean(measurement),
    median = median(measurement),
    min = min(measurement),
    max = max(measurement),
```

```
sd = sd(measurement),  
iqr = IQR(measurement)  
)
```