

# Homework 1

This homework is due on the deadline posted on edX. Please submit a .pdf file of your output and upload a .zip file containing your .Rmd file. Do NOT include your name or EID in your filenames.

In this homework you will be working with the `chickwts` dataset built into R. This data set contains weight measurements of chicks fed on different food sources to assess their effectiveness on growth rate.

```
head(chickwts)
```

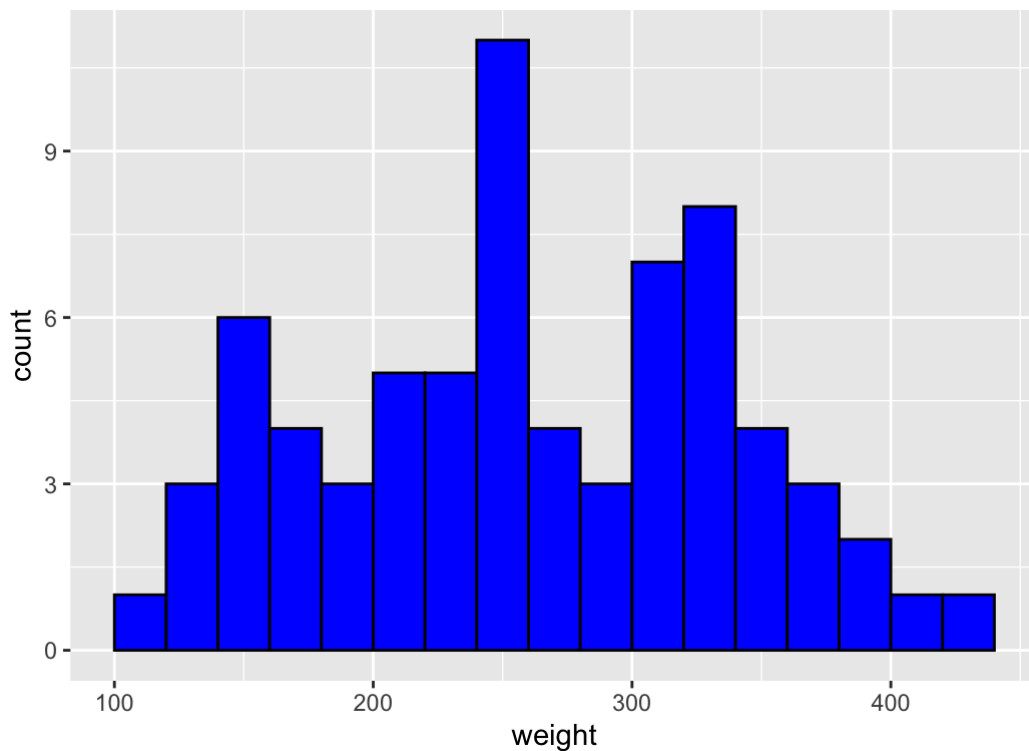
```
##    weight      feed
## 1    179 horsebean
## 2    160 horsebean
## 3    136 horsebean
## 4    227 horsebean
## 5    217 horsebean
## 6    168 horsebean
```

**Problem 1:** Use ggplot to make a histogram of the `weight` column. Manually choose appropriate values for `binwidth` and `center`. Explain your choice of values in 2-3 sentences.

```
summary(chickwts)
```

```
##      weight      feed
## Min.   :108.0 casein   :12
## 1st Qu.:204.5 horsebean:10
## Median :258.0 linseed  :12
## Mean   :261.3 meatmeal :11
## 3rd Qu.:323.5 soybean  :14
## Max.   :423.0 sunflower:12
```

```
ggplot(chickwts, aes(weight)) +
  geom_histogram(
    binwidth = 20,
    center = 10,
    fill = 'blue',
    color = 'black')
```



*Your explanation goes here.*

Before selecting the *binwidth* and *center*, I decided to look at the summary statistics of the dataset, which would help in evaluating my choices. Upon noticing that the weights range from ~100 to ~420, I experimented with different multiples of 10, and found that using **20** as my binwidth provided interesting insights - there are several peaks and troughs amongst the distribution of weights. Finally, I selected my center as **10** because it presents bins in a clean fashion (i.e. from 100 to 120, 200 to 220, etc.).

**Problem 2:** Modify the plot from Problem 1 to show one panel per feed. Hint: Use `facet_wrap()` .

```
ggplot(chickwts, aes(weight)) +  
  geom_histogram(  
    binwidth = 20,  
    center = 10,  
    fill = 'blue',  
    color = 'black') +  
  facet_wrap(vars(feed))
```

