WU #4 - Sums of Squares

Math 158 - Jo Hardin

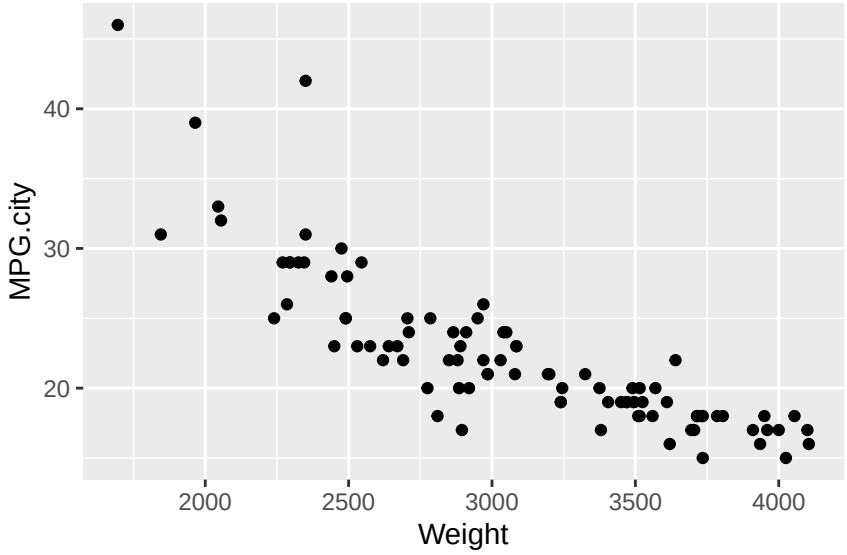
Tuesday 2/1/2022

Name: _____

Names of people you worked with: _____

Consider the following ANOVA table. The data are based on a random sample of cars from among 1993 passenger car models that were listed in both *Consumer Reports* and the *PACE Buying Guide*. We are considering the variables `weight` and `MPG.city`.

```
Cars93 %>%
  ggplot(aes(y=MPG.city, x= Weight)) +
  geom_point()
```




```
Cars93 %>%
  lm(MPG.city ~ Weight, data = .) %>%
  anova() %>%
  tidy()
```

```
## # A tibble: 2 x 6
##   term         df sumsq  meansq statistic   p.value
##   <chr>     <int> <dbl>   <dbl>     <dbl>     <dbl>
## 1 Weight        1 2066. 2066.        224.  2.97e-26
## 2 Residuals    91  840.    9.23       NA  NA
```

1. Find R^2.
2. Interpret R^2.

Solution:

1.

```
Cars93 %>%
  lm(MPG.city ~ Weight, data = .) %>%
  anova() %>%
  tidy() %>%
  mutate(R2 = 1 - sumsq / sum(sumsq)) %>%
  filter(term == "Residuals") %>%
  dplyr::select(R2)
```

```
## # A tibble: 1 x 1
##      R2
##   <dbl>
## 1 0.711
```

Check:

```
Cars93 %>%
  lm(MPG.city ~ Weight, data = .) %>%
  glance()
```

```
## # A tibble: 1 x 12
##   r.squared adj.r.squared sigma statistic  p.value    df logLik
##       <dbl>         <dbl> <dbl>     <dbl>    <dbl> <dbl>  <dbl>
## 1     0.711         0.708  3.04      224. 2.97e-26     1  -234.
## # ... with 5 more variables: AIC <dbl>, BIC <dbl>,
## #   deviance <dbl>, df.residual <int>, nobs <int>
```

2. 71.1% of the variability in the `MPG` that a car gets in the city can be explained by the linear model with `weight` of the car.