Stat 140 - Chapters 7 and 8 Quiz

What's Your Name? ______________________________

Which section are you in? ______________________________

You may refer to any old homework assignments and any materials on the course website, but you must work on this alone.

Old Faithful

Old Faithful is a geyser in Yellowstone National Park in Wyoming, US. The geyser periodically erupts, sending a column of hot water 100 to 200 feet into the air for about 1 to 5 minutes per eruption. The data set we'll work with in this quiz has information about all eruptions that occurred between August 1, 1985 and August 15, 1985. Specifically, for each eruption in that time span we have recorded the duration of the eruption in minutes (`eruption_duration_min`) and the time elapsed in minutes between that eruption and the next eruption after it (`time_to_next_eruption_min`). The following output from R shows the first few rows and structure of the data set, a plot of `eruption_duration_min` vs. `time_to_next_eruption_min`, and output from a linear model fit for these data. Use this R output to answer the questions below.

```r
head(old_faithful)
```

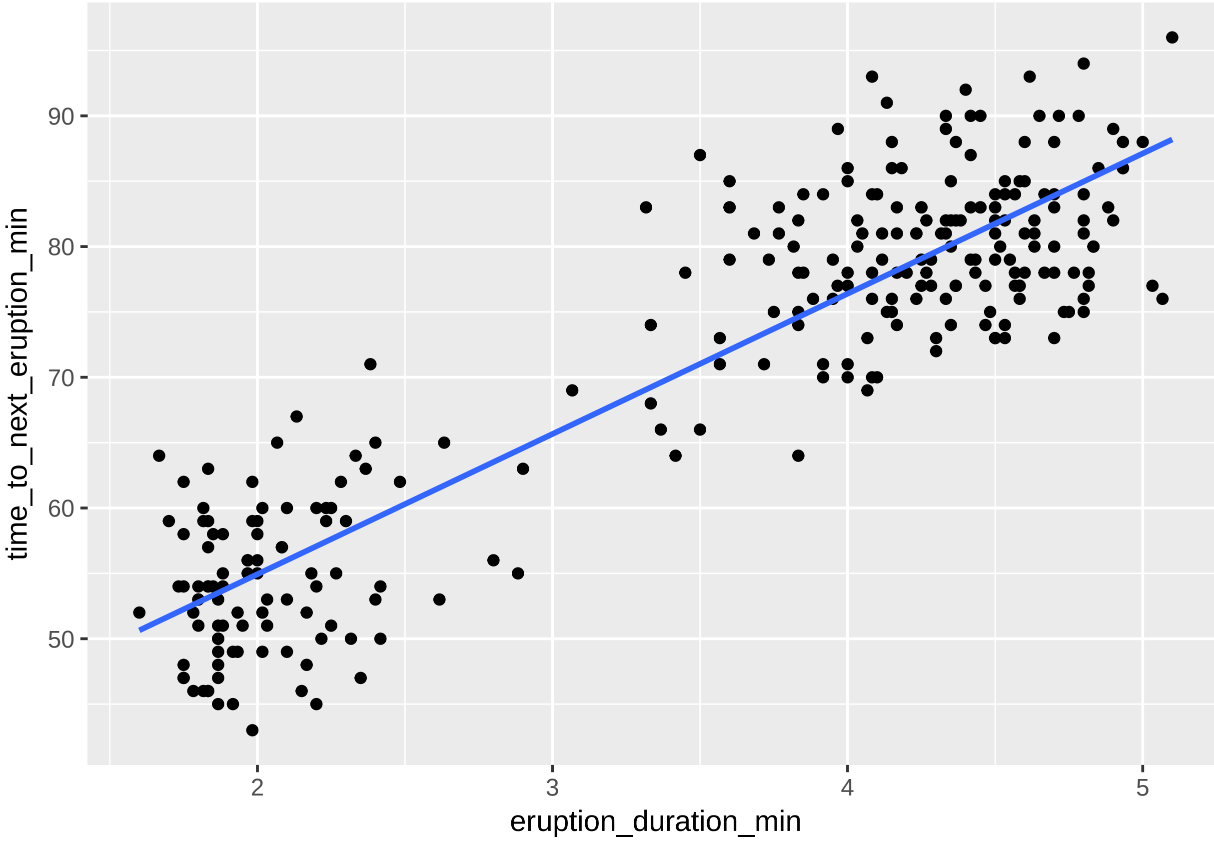
```
## # A tibble: 6 x 2
##   eruption_duration_min time_to_next_eruption_min
##                   <dbl>                     <int>
## 1                 3.600                        79
## 2                 1.800                        54
## 3                 3.333                        74
## 4                 2.283                        62
## 5                 4.533                        85
## 6                 2.883                        55
```

```r
str(old_faithful)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    272 obs. of  2 variables:
##  $ eruption_duration_min    : num  3.6 1.8 3.33 2.28 4.53 ...
##  $ time_to_next_eruption_min: int  79 54 74 62 85 55 88 85 51 85 ...
##  - attr(*, "spec")=List of 2
##   ..$ cols   :List of 2
##   .. ..$ eruption_duration_min    : list()
##   .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
##   .. ..$ time_to_next_eruption_min: list()
##   .. .. ..- attr(*, "class")= chr  "collector_integer" "collector"
##   ..$ default: list()
##   .. ..- attr(*, "class")= chr  "collector_guess" "collector"
##   ..- attr(*, "class")= chr "col_spec"
```

```r
ggplot() +
  geom_point(mapping = aes(x = eruption_duration_min, y = time_to_next_eruption_min),
    data = old_faithful) +
  geom_smooth(mapping = aes(x = eruption_duration_min, y = time_to_next_eruption_min),
    data = old_faithful,
```

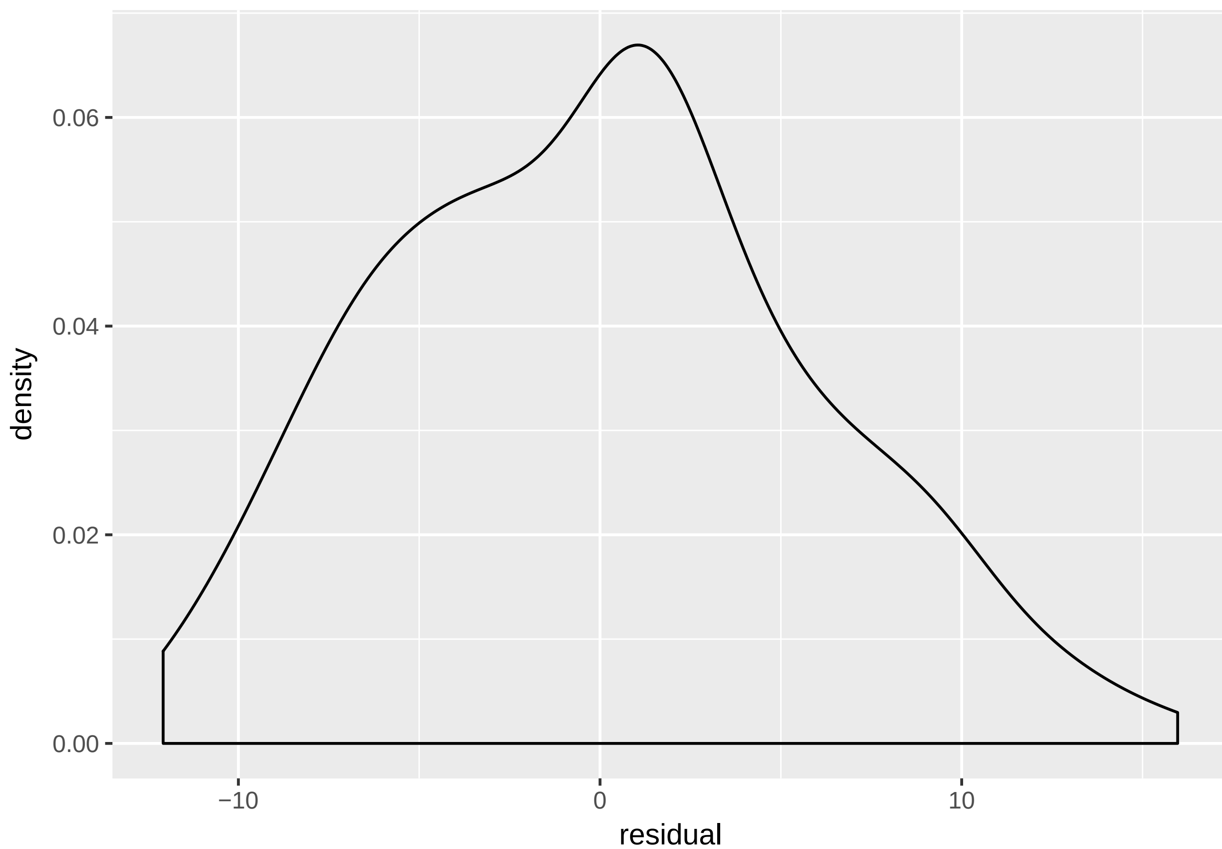
```r
    method = "lm",
    se = FALSE)
```




```r
linear_fit <- lm(time_to_next_eruption_min ~ eruption_duration_min,
  data = old_faithful)

old_faithful <- mutate(old_faithful,
  residual = residuals(linear_fit),
  predicted = predict(linear_fit)
)

ggplot() +
  geom_density(mapping = aes(x = residual), data = old_faithful)
```




```r
coef(linear_fit)
```

```
##          (Intercept) eruption_duration_min
##             33.47440              10.72964
```

```r
summary(linear_fit)
```

```
##
## Call:
## lm(formula = time_to_next_eruption_min ~ eruption_duration_min,
##     data = old_faithful)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.0796  -4.4831   0.2122   3.9246  15.9719
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            33.4744     1.1549   28.98   <2e-16 ***
## eruption_duration_min  10.7296     0.3148   34.09   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.914 on 270 degrees of freedom
## Multiple R-squared:  0.8115, Adjusted R-squared:  0.8108
## F-statistic:  1162 on 1 and 270 DF,  p-value: < 2.2e-16
```

1. Based on the scatter plot of the data at the top of page 2 and the density plot at the top of page 3, is a linear model appropriate for modeling the relationship between the eruption duration and the waiting time to the next eruption? Discuss all assumptions. For each assumption, indicate which plot is relevant to checking that assumption and whether the assumption appears to be satisfied based on the plot.

2. Write down the linear model equation to obtain the predicted waiting time for the next eruption if I know how long the most recent eruption lasted.

3. Interpret the intercept and slope in the context of this example.

4. Suppose I visit Yellowstone National Park, and when I arrive the park ranger tells me that the Old Faithful geyser *just* stopped erupting as I drove in. The ranger says that the eruption that just finished lasted for 4.5 minutes. How long does the model predict that I will have to wait until the next eruption?

5. Find the model's prediction for the waiting time until the next eruption if the most recent eruption lasted 30 minutes. Would you expect the prediction of waiting time for the next eruption from this linear model to be useful? Why or why not?

6. What is the residual standard deviation? Use the "95" part of the 68-95-99.7 rule to interpret the residual standard deviation in the context of this problem, including the units.

7. What is the R^2? Interpret the R^2 in the context of this problem.