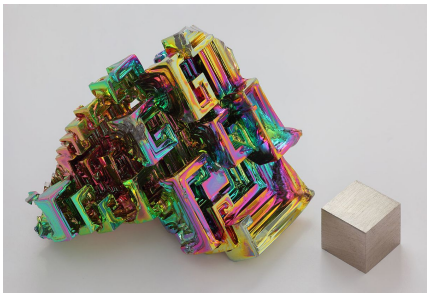
Stat 242 Quiz – Topics Drawn from Chapter 8

What's Your Name? _____

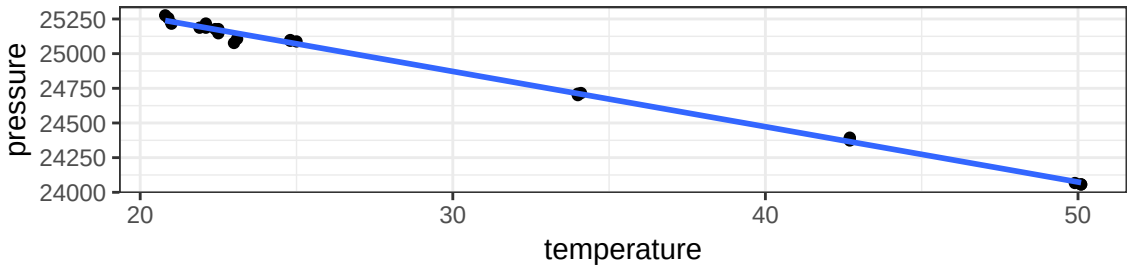


Bismuth is a metal that forms a crystal structure, as illustrated above. The crystal structure of bismuth changes depending on the temperature and pressure it is subjected to; there are four different crystal structures that will be realized at different temperatures and pressures. Houck studied the relationship between the temperature of Bismuth and the pressure at which it undergoes a first change in crystal structure, from Bi-I to Bi-II (Houck, J.C. (1970). Temperature coefficient of the bismuth I-II transition pressure. *J. Res.Nat. Bur. Stand.*, 74 A, 51-54). In this experiment, each sample was maintained at a specified temperature while the pressure was adjusted. The pressure at which each sample changed crystal structure was recorded.

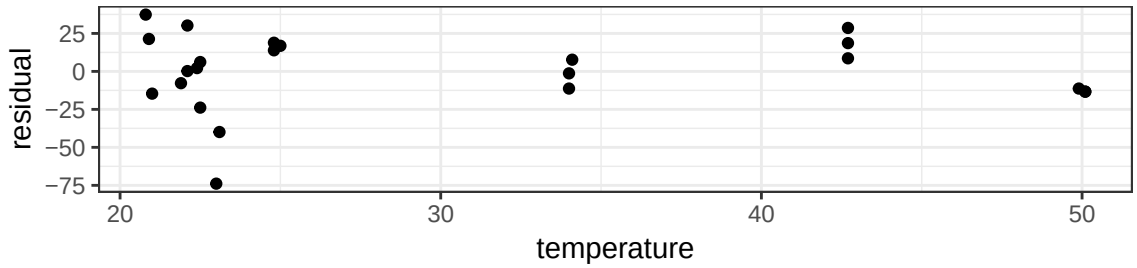
In this experiment, the researchers had experimental equipment that could be set for a specified temperature. To collect the data, they set this equipment to a specified temperature, and then ran several samples at that temperature before setting the equipment to a different temperature.

Here is a scatter plot of the data with a regression line fit overlaid on top, a scatter plot of residuals vs. temperature, and a histogram of the residuals.

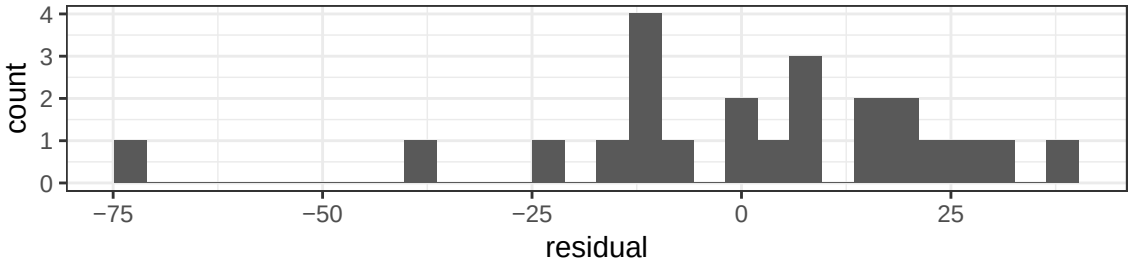
```
ggplot(data = bismuth, mapping = aes(x = temperature, y = pressure)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  theme_bw()
```




```
ggplot(data = bismuth, mapping = aes(x = temperature, y = residual)) +
  geom_point() +
  theme_bw()
```




```
ggplot(data = bismuth, mapping = aes(x = residual)) +
  geom_histogram() +
  theme_bw()
```



1. Check each of the regression model conditions. For each condition, write a sentence or two describing whether or not the condition is satisfied and why. If your conclusion is based on the plots above, please clearly indicate which plot or plots you are looking at and describe a specific characteristic of that plot that your conclusion is based on. For any conditions that are not satisfied, suggest a possible strategy for addressing the problem.

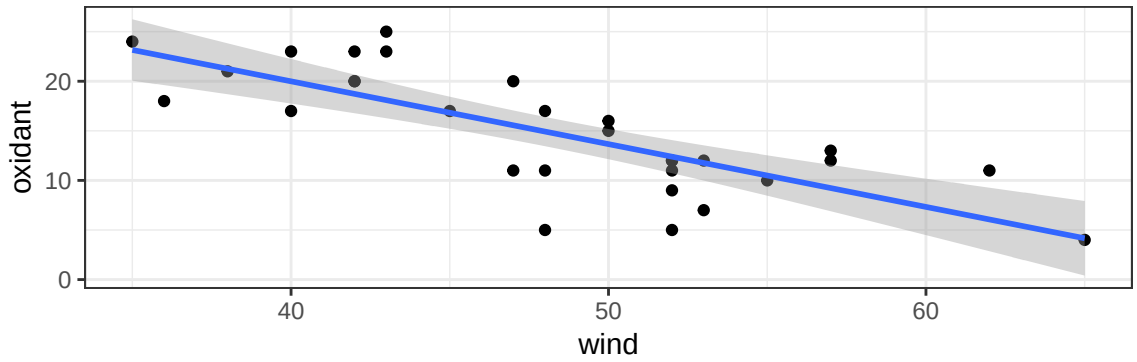
Stat 242 Quiz – Topics Drawn from Chapter 8

What's Your Name? _____

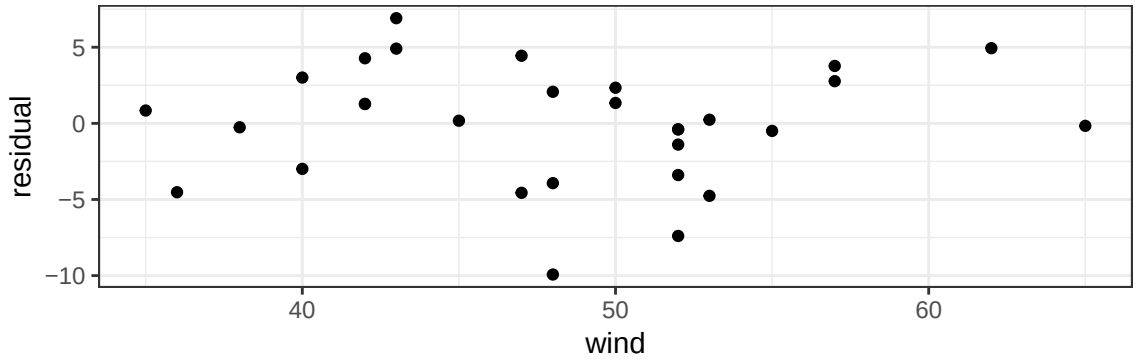
The Los Angeles Pollution Control District records the levels of pollutants and various meteorological conditions, and attempts to construct models to predict pollution levels and to gain a better understanding of the complexities of air pollution. Here we will consider a small subset of this data, with measurements of the maximum level of an oxidant (a photochemical pollutant) and average wind speeds in the mornings of 50 days during one summer.

Here is a scatter plot of the data with a regression line fit overlaid on top, a scatter plot of residuals vs. wind speeds, and a histogram of the residuals.

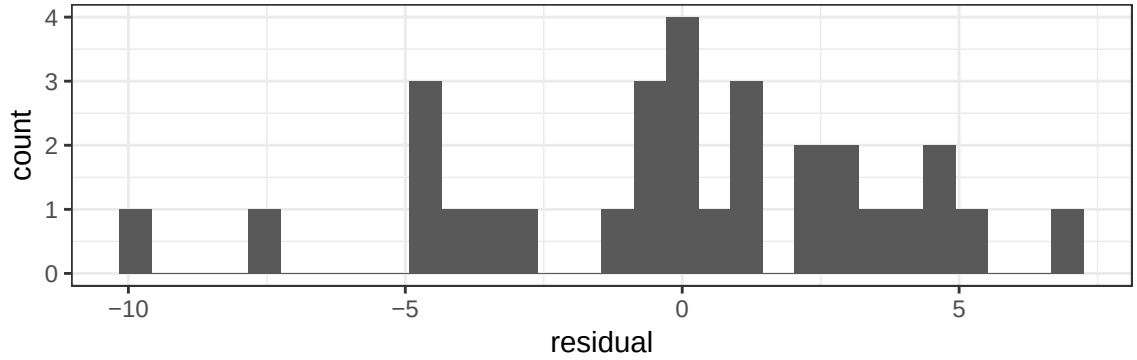
```
ggplot(data = air_pollution, mapping = aes(x = wind, y = oxidant)) +
  geom_point() +
  geom_smooth(method = "lm") +
  theme_bw()
```




```
ggplot(data = air_pollution, mapping = aes(x = wind, y = residual)) +
  geom_point() +
  theme_bw()
```




```
ggplot(data = air_pollution, mapping = aes(x = residual)) +
  geom_histogram() +
  theme_bw()
```



1. Check each of the regression model conditions. For each condition, write a sentence or two describing whether or not the condition is satisfied and why. If your conclusion is based on the plots above, please clearly indicate which plot or plots you are looking at and describe a specific characteristic of that plot that your conclusion is based on. For any conditions that are not satisfied, suggest a possible strategy for addressing the problem.

Stat 242 Quiz – Topics Drawn from Chapter 8

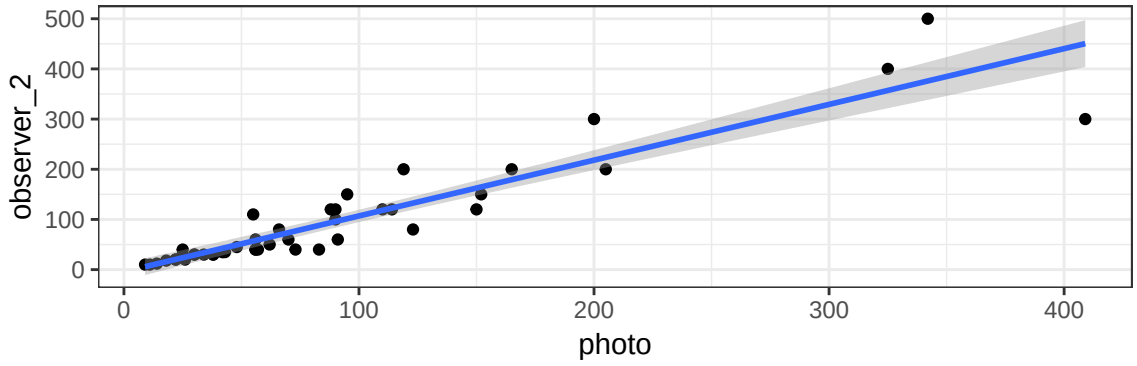
What's Your Name? _____

Aerial survey methods are sued to estimate the number of snow geese in their summer range areas westof Hudson's Bay in Canada. To obtain estimates, small aircraft fly over the range and, when a flock of snow geese is spotted, an experienced observed estimates teh number of geese in teh flock. To investigate the reliability of this method, an experiment in which an airplane carried two observers flew over 45 flocks, and each observer independently estimated the number of geese in the flock. Also, a photograph of the flock was taken so that an exact count of the number of geese in the flock could be obtained (Weisberg, 1985).

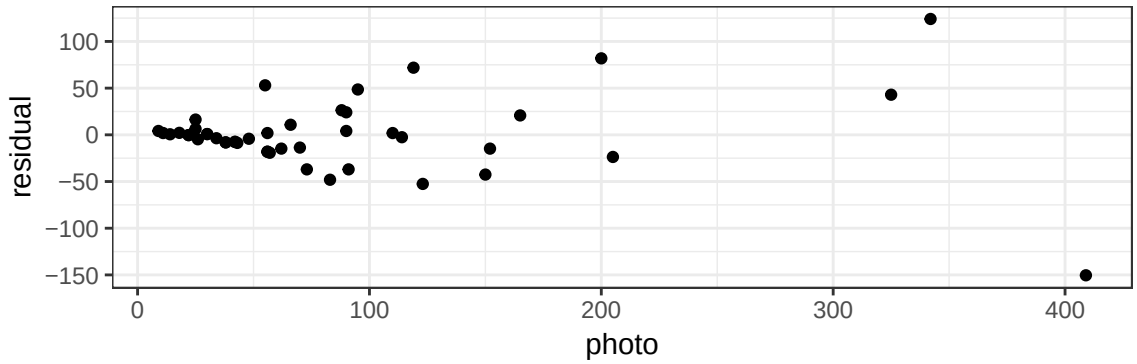
Here we will just look at the estimated counts from the second observer (our response variable), and see how they compare to the exact counts from the photograph (explanatory variable).

Here is a scatter plot of the data with a regression line fit overlaid on top, a scatter plot of residuals vs. the exact count from the photo, and a histogram of the residuals.

```
ggplot(data = geese, mapping = aes(x = photo, y = observer_2)) +
  geom_point() +
  geom_smooth(method = "lm") +
  theme_bw()
```

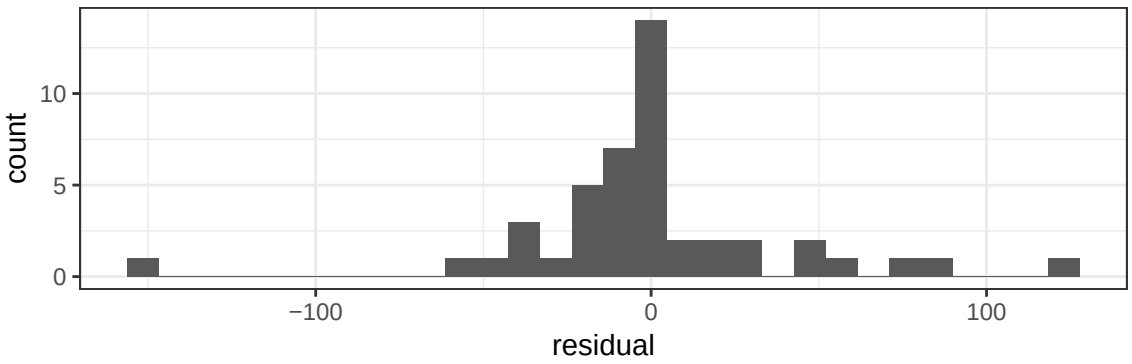



```
ggplot(data = geese, mapping = aes(x = photo, y = residual)) +
  geom_point() +
  theme_bw()
```




```
ggplot(data = geese, mapping = aes(x = residual)) +
  geom_histogram() +
  theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



1. Check each of the regression model conditions. For each condition, write a sentence or two describing whether or not the condition is satisfied and why. If your conclusion is based on the plots above, please clearly indicate which plot or plots you are looking at and describe a specific characteristic of that plot that your conclusion is based on. For any conditions that are not satisfied, suggest a possible strategy for addressing the problem.

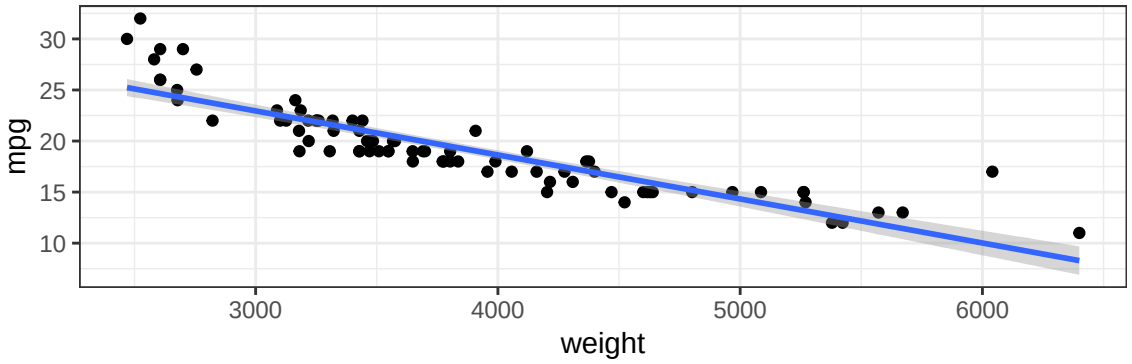
Stat 242 Quiz – Topics Drawn from Chapter 8

What's Your Name? _____

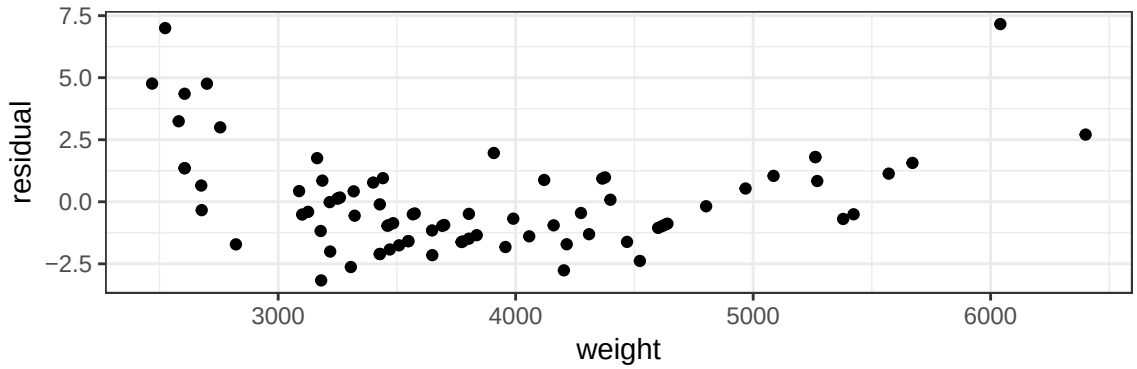
We know that heavier cars need more fuel, but exactly how does a car's weight affect its fuel efficiency? We have a data set with the weight (the explanatory variable) and fuel efficiency (the response variable, in miles per gallon, mpg) for 80 cars.

Here is a scatter plot of the data with a regression line fit overlaid on top, a scatter plot of residuals vs. the exact count from the photo, and a histogram of the residuals.

```
ggplot(data = cars, mapping = aes(x = weight, y = mpg)) +
  geom_point() +
  geom_smooth(method = "lm") +
  theme_bw()
```

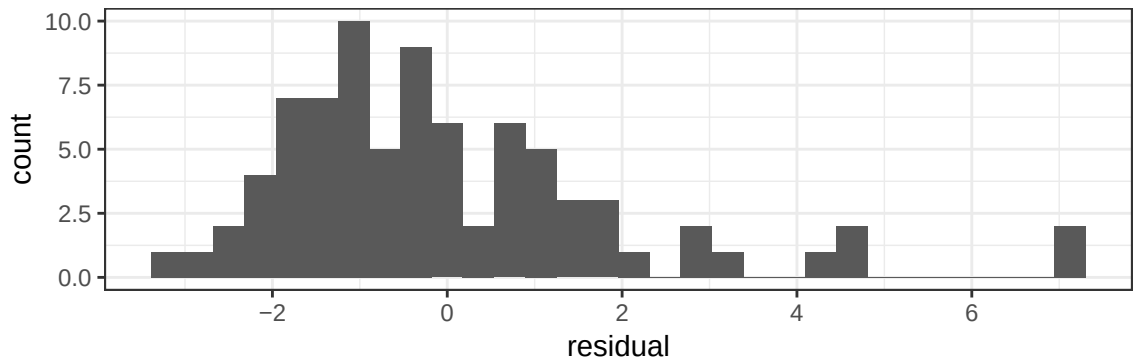



```
ggplot(data = cars, mapping = aes(x = weight, y = residual)) +
  geom_point() +
  theme_bw()
```




```
ggplot(data = cars, mapping = aes(x = residual)) +
  geom_histogram() +
  theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



1. Check each of the regression model conditions. For each condition, write a sentence or two describing whether or not the condition is satisfied and why. If your conclusion is based on the plots above, please clearly indicate which plot or plots you are looking at and describe a specific characteristic of that plot that your conclusion is based on. For any conditions that are not satisfied, suggest a possible strategy for addressing the problem.