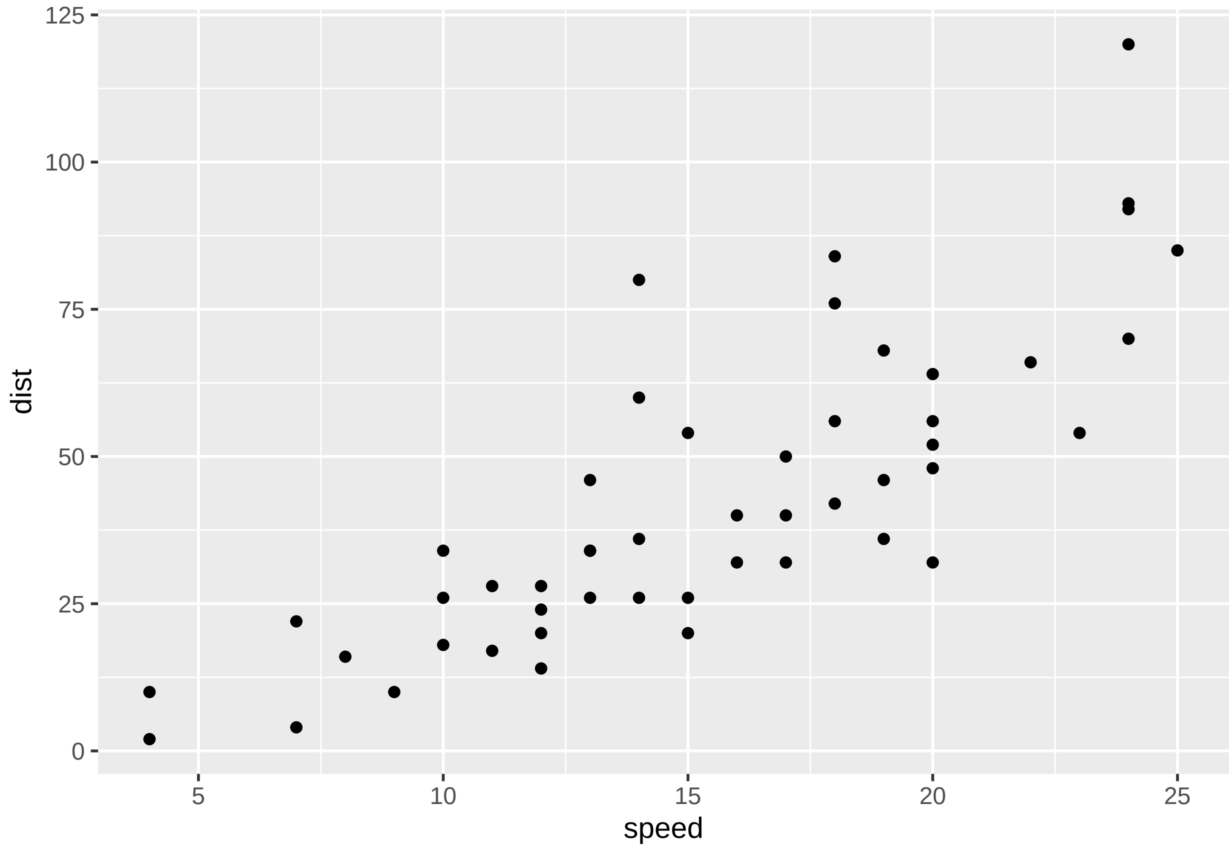
# Module 3 Lab Submission

Nora Quick

Consider the `cars` data, which contains cars speed in MPH and stopping distance in feet. Load the data with `data("cars")`.

```
data("cars")
qplot(speed, dist, data = cars)
```




- Fit a simple linear model with `dist` as the response and `speed` as the explanatory variable.

```
fit <- lm(dist ~ speed, data = cars)
summary(fit)
```

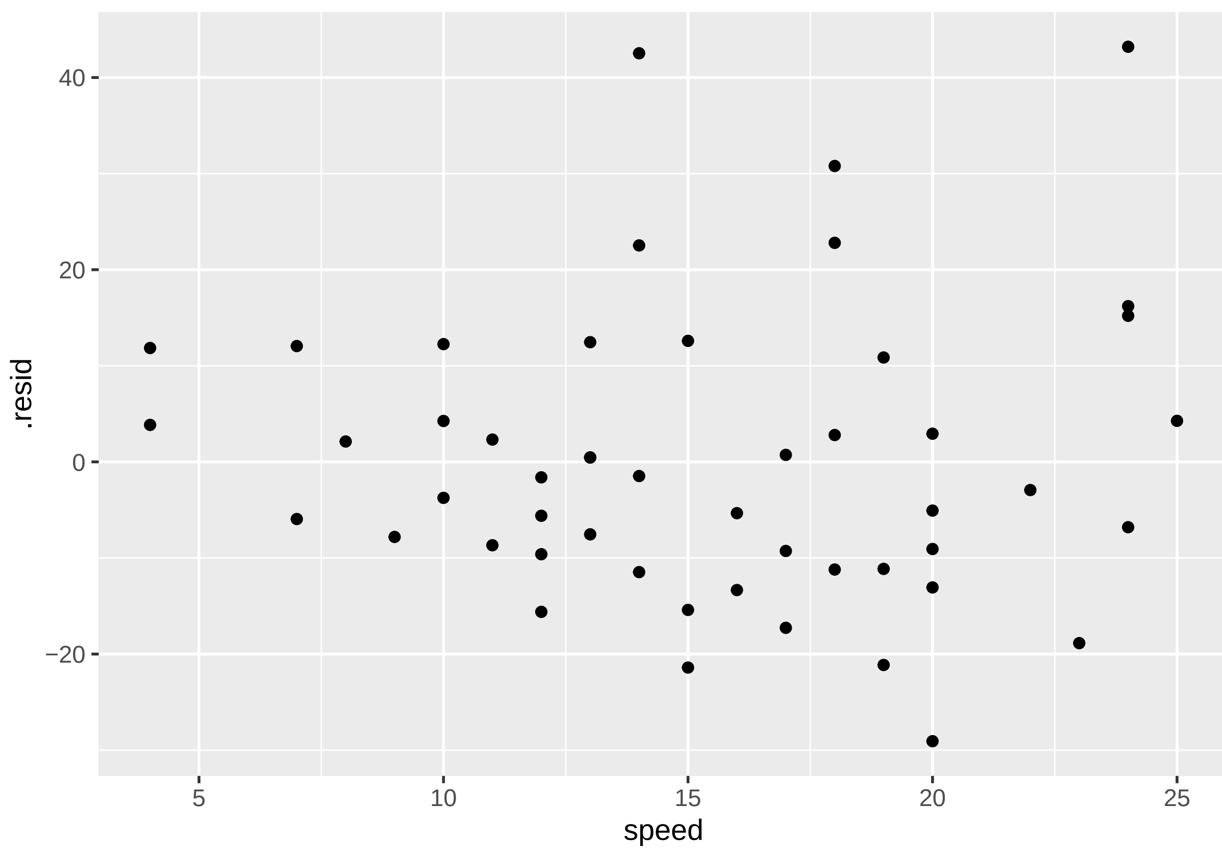
```
##
## Call:
## lm(formula = dist ~ speed, data = cars)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.069  -9.525  -2.272   9.215  43.201
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.5791     6.7584  -2.601   0.0123 *
## speed         3.9324     0.4155   9.464 1.49e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.38 on 48 degrees of freedom
## Multiple R-squared:  0.6511, Adjusted R-squared:  0.6438
## F-statistic: 89.57 on 1 and 48 DF,  p-value: 1.49e-12
```

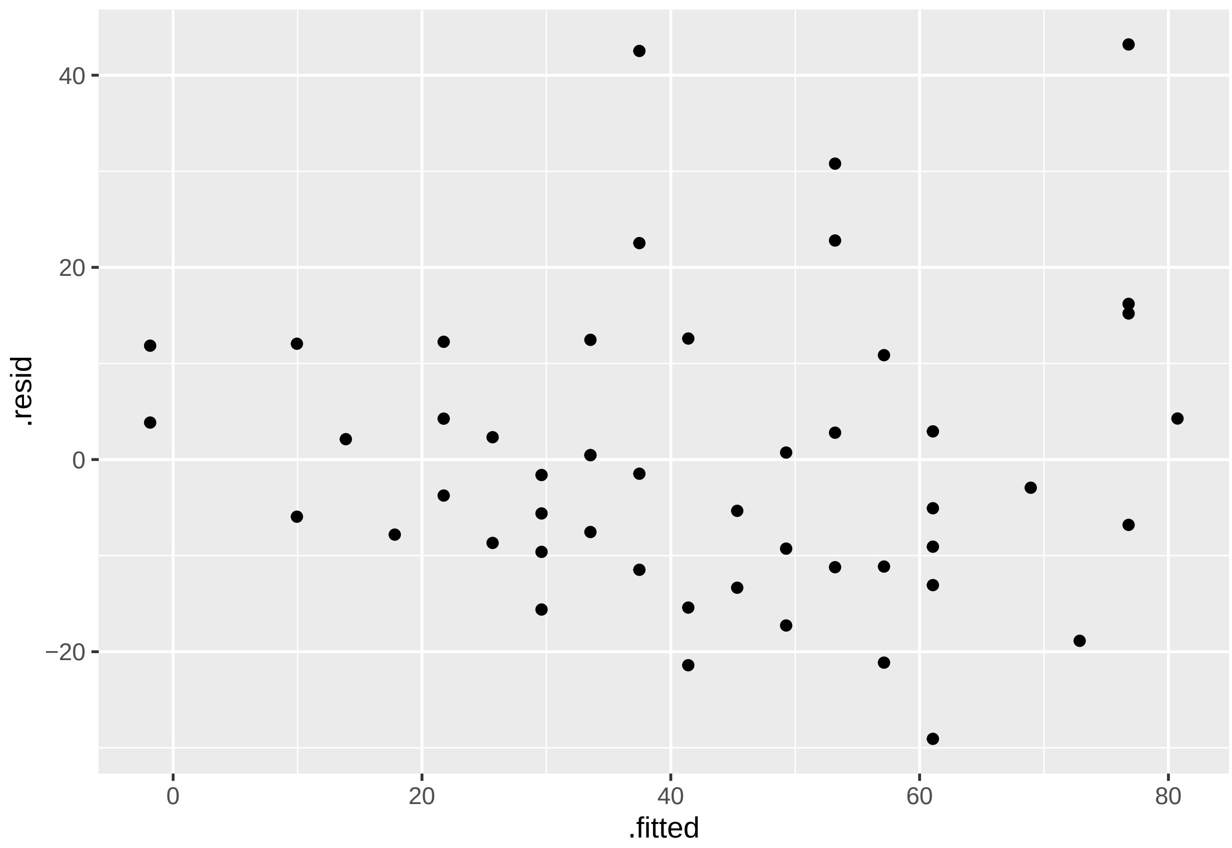
- Create two diagnostic plots using the residuals, one with `speed` on the x-axis, and the other with the fitted values from the model. Do the plots look good: do these data seem to satisfy the assumptions for a linear regression model?

```
f <- augment(fit)

qplot(speed, .resid, data = f)
```



```r
qplot(.fitted, .resid, data = f)
```




- Use `predict()` to get the confidence and prediction intervals using the following new data.

```r
new <- data.frame(speed = c(6, 10.5, 14.7, 18.3, 21))

predict(fit, newdata = new, interval = "confidence")
```

```
##         fit       lwr      upr
## 1  6.015358 -2.973341 15.00406
## 2 23.711197 17.720996 29.70140
## 3 40.227314 35.815250 44.63938
## 4 54.383985 49.384564 59.38341
## 5 65.001489 58.597384 71.40559
```

Now note that there are many speeds for which there were multiple observations at that speed. This means we can perform a lack-of-fit test on this data.

- Fit a separate means model using `lm()` and `factor()` to treat `speed` as a categorical variable.

```r
lm_speed <- lm(dist ~ speed, data = cars)
lm_speed
```

```
##
## Call:
## lm(formula = dist ~ speed, data = cars)
##
## Coefficients:
## (Intercept)        speed
##     -17.579        3.932
```

```r
fac_speed <- lm(dist ~ factor(speed), data = cars)
fac_speed
```

```
##
## Call:
## lm(formula = dist ~ factor(speed), data = cars)
##
## Coefficients:
##     (Intercept)   factor(speed)7   factor(speed)8   factor(speed)9
##            6.00             7.00            10.00             4.00
## factor(speed)10  factor(speed)11  factor(speed)12  factor(speed)13
##           20.00            16.50            15.50            29.00
## factor(speed)14  factor(speed)15  factor(speed)16  factor(speed)17
##           44.50            27.33            30.00            34.67
## factor(speed)18  factor(speed)19  factor(speed)20  factor(speed)22
##           58.50            44.00            44.40            60.00
## factor(speed)23  factor(speed)24  factor(speed)25
##           48.00            87.75            79.00
```

- Compare the separate means model to the simple linear regression model using the `anova()` function.

```r
anova(lm_speed, fac_speed)
```

```
## Analysis of Variance Table
##
## Model 1: dist ~ speed
## Model 2: dist ~ factor(speed)
##   Res.Df     RSS Df Sum of Sq      F Pr(>F)
## 1     48 11353.5
## 2     31  6764.8 17    4588.7 1.2369 0.2948
```