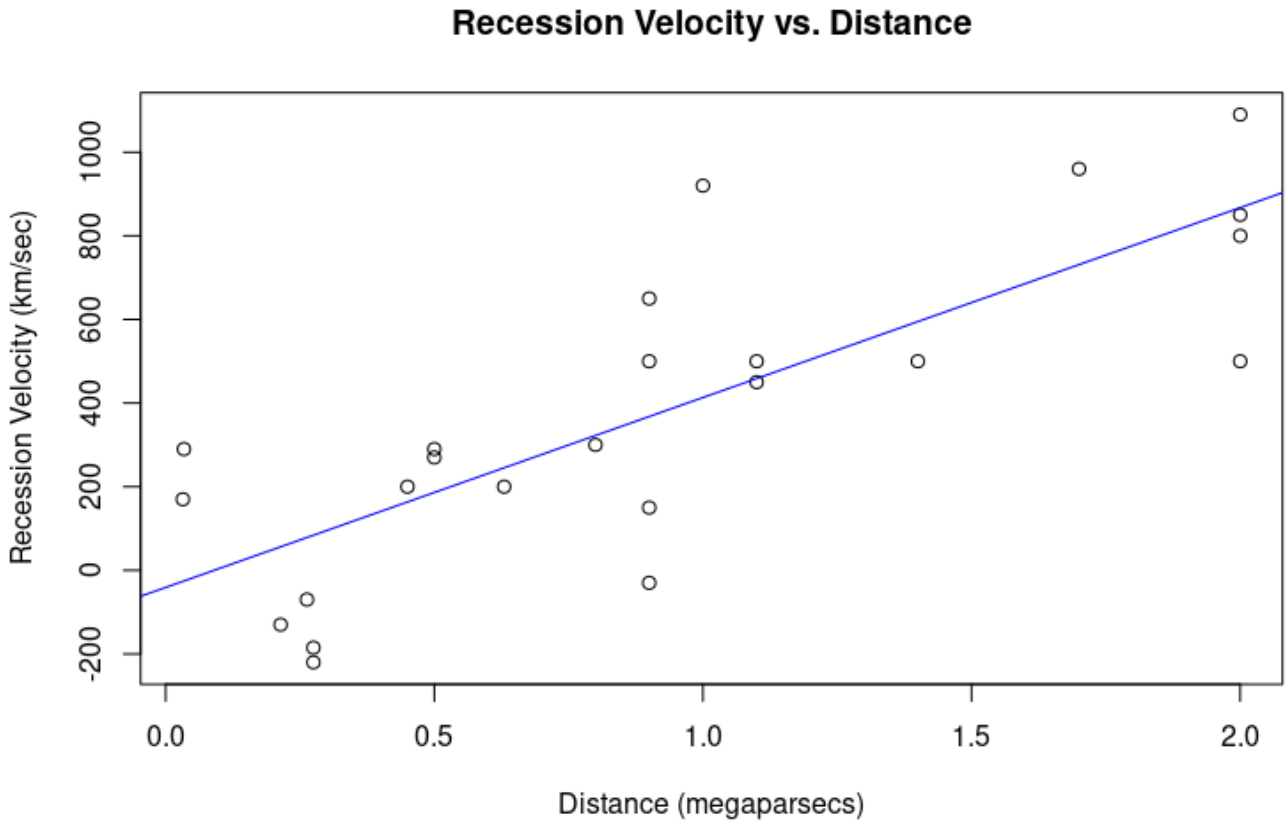
Assignment 2

STAT463

1) Visually display the data in an appropriate graph and comment on anything that may be of note. Be sure to adequately label the graph, indicating in a title what is being displayed as well as what each axis represent (including units of measurement).



There is a loose correlation between recession velocity and distance. As distance increases, so does recession velocity.

2. It is desired to ascertain whether galaxies that are farther away tend to be receding faster from Earth (i.e. they have higher recession velocities). Based on the plot you constructed in part (1), would a linear regression model with recession velocity as the response variable and distance as the explanatory variable be appropriate for this dataset?

Based on the plot constructed, a linear regression model with distance as the explanatory variable and recession velocity as the response variable is appropriate. There is a visible linear correlation between distance and recession velocity. Furthermore, we want to determine if recession velocity correlates to distance.

3. What is the nature of the relationship between galaxies' distances from earth and their recession velocities? Specifically, as the distance increases, how is the recession velocity affected? Quantify it and interpret this quantity.

As distance increases, so does recession velocity. If we use a linear regression model, we have the equation, $f(x) = -40.78365 + 454.15844x$ to model the relationship between distance and recession velocity.

4. Does the data suggest that there is a statistically significant relationship between galaxies' recession velocities and their distances from earth?

The data does suggest that there is a statistically significant relationship between galaxies' recession velocity and their distances from earth. We can perform an F-test which yields a P-value of 0.000004477491, which is extremely small, and therefore, statistically significant.

5. How well do galaxies' distances from earth explain their recession velocities? Quantify this and interpret this quantity.

If we compute the correlation coefficient, we get that $r^2 = 0.6235$, showing that there is a strong linear relationship. However, it does not fully explain the recession velocity.

6. Andromeda, the nearest large spiral galaxy to our home galaxy the Milky Way, is 0.77 megaparsecs away. Give an estimate of its recession velocity. Additionally, provide a range of value that will encompass Andromeda's true recession velocity with probability 0.95.

Using our regression line equation from earlier, $f(x) = -40.78365 + 454.15844x$, the estimated recession velocity would be $308.9183488\ \mathrm{km/sec}$.

A range of values that encompasses Andromeda's true recession velocity with probability 0.95, is $(-184.5625, 802.3992)$