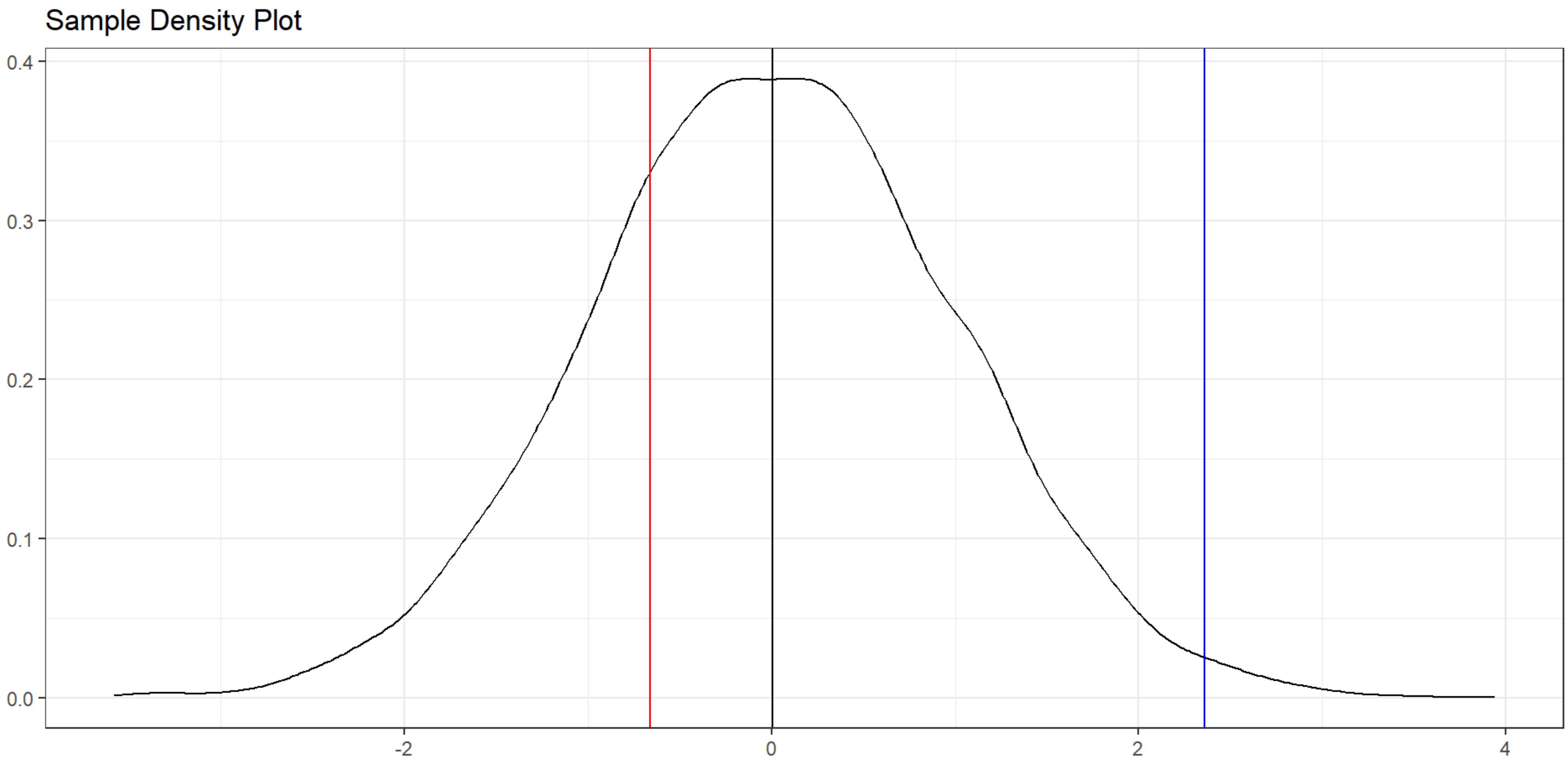
Quantile Regression

Chris Cioffi, Kristina Frazier, Aidan Hennessy, Mike McHenry

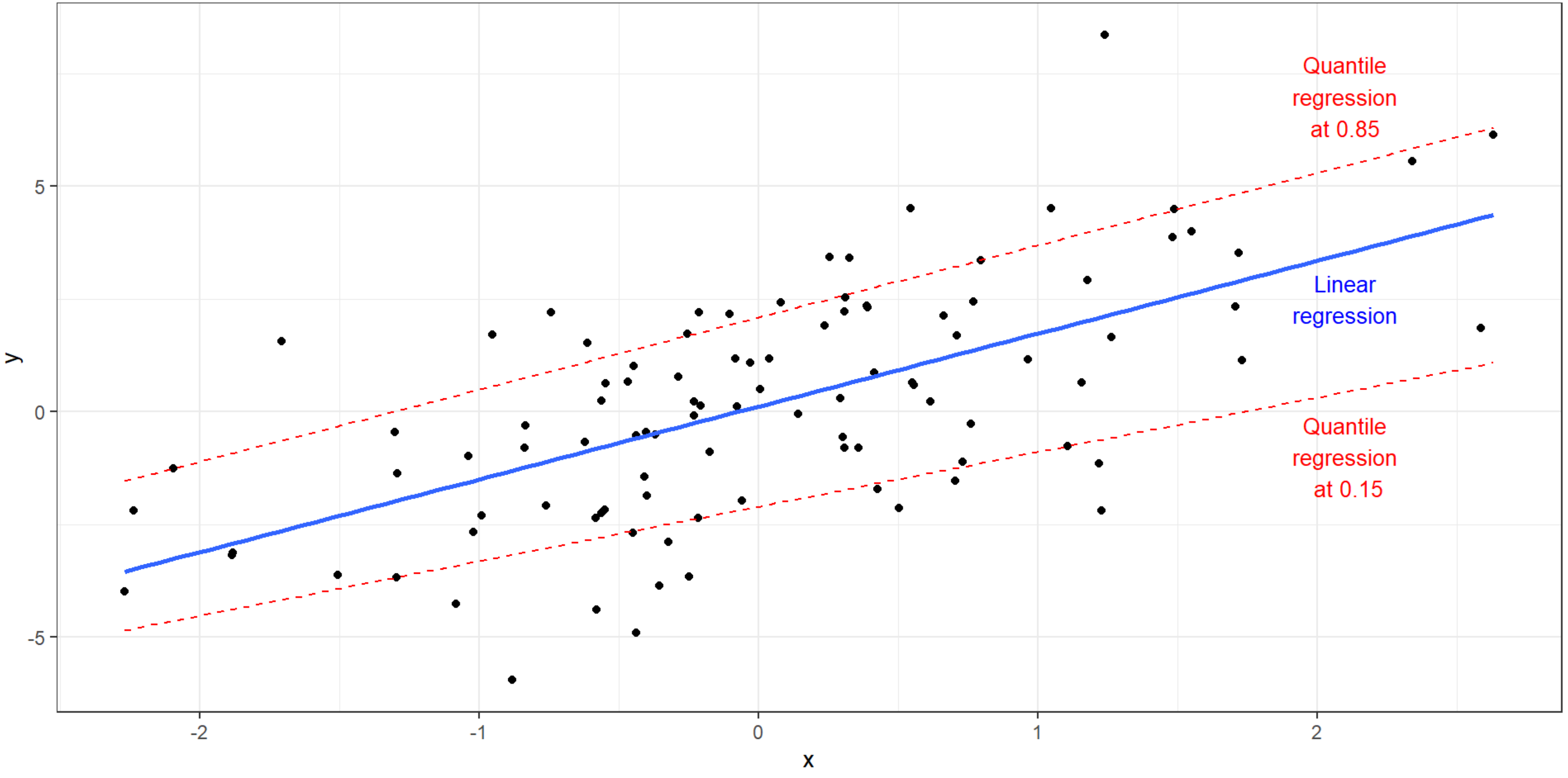
Quantile Regression

- A useful alternative to Ordinary Least Squares (OLS).
- Produces estimated regression coefficients for specific quantiles of the response variable.
    - As a result, the coefficients will change as the specified quantile of the response variable changes.

Sample Density



Example of Linear and Quantile Trend Lines



Why Use Quantile Regression?

- The relationship between the predictor and response variables may vary across the distribution, making it useful to allow the estimates to change.
- More robust to outliers.
- If OLS assumptions are violated.
    - Nonconstant variance, nonnormality, indpendence, nonlinearity.

General Form of the Model

- $Q_\tau(Y) = \beta_0(\tau) + \beta_1(\tau)X_1 + \ldots + \beta_p(\tau)X_p$
- Where τ is a specific quantile value.
- The objective of quantile regression is to minimize the median absolute deviations (Quantile Loss).
- Rather than minimizing the sum of squares in OLS.

Model Assumptions

- Large sample size.
- Response variable is continuous.
- No strict distribution to adhere to.

Quantreg

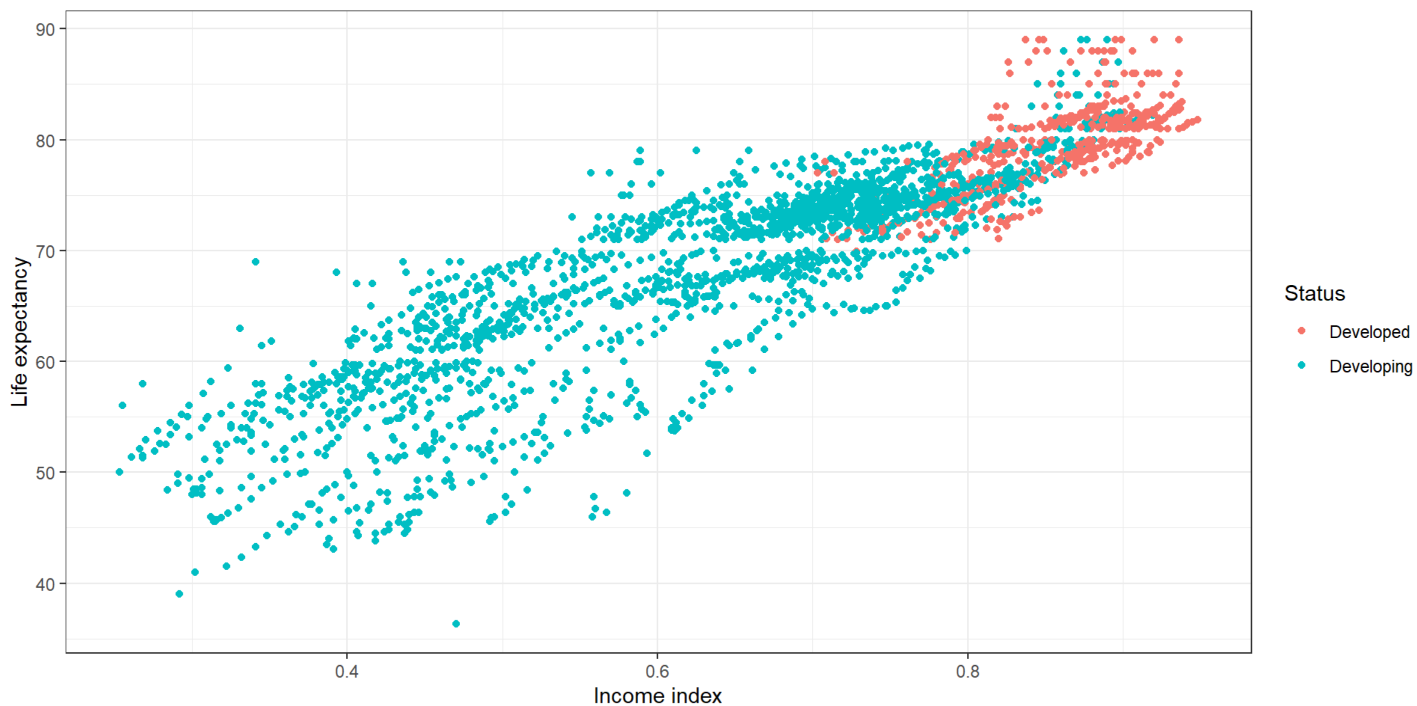
- While the lm() function allows us to use linear regression and estimate the conditional mean of the response, considering explanatory variable values, quantreg can be used to easily conduct quantile regresion in R.

Quantreg and WHO data

- Demo: estimate the conditional quantiles, considering the predictors, and here we can use quantreg to compare world nations' life expectancies to its so-called income index.
- The dataset is maintained by the World Health Organization and looks at a variety of public health and income factors for world nations. We're using a Kaggle version, which can be easily downloaded here.
- *Demo adapted from examples found on Medium and R-bloggers

Quantreg and WHO data

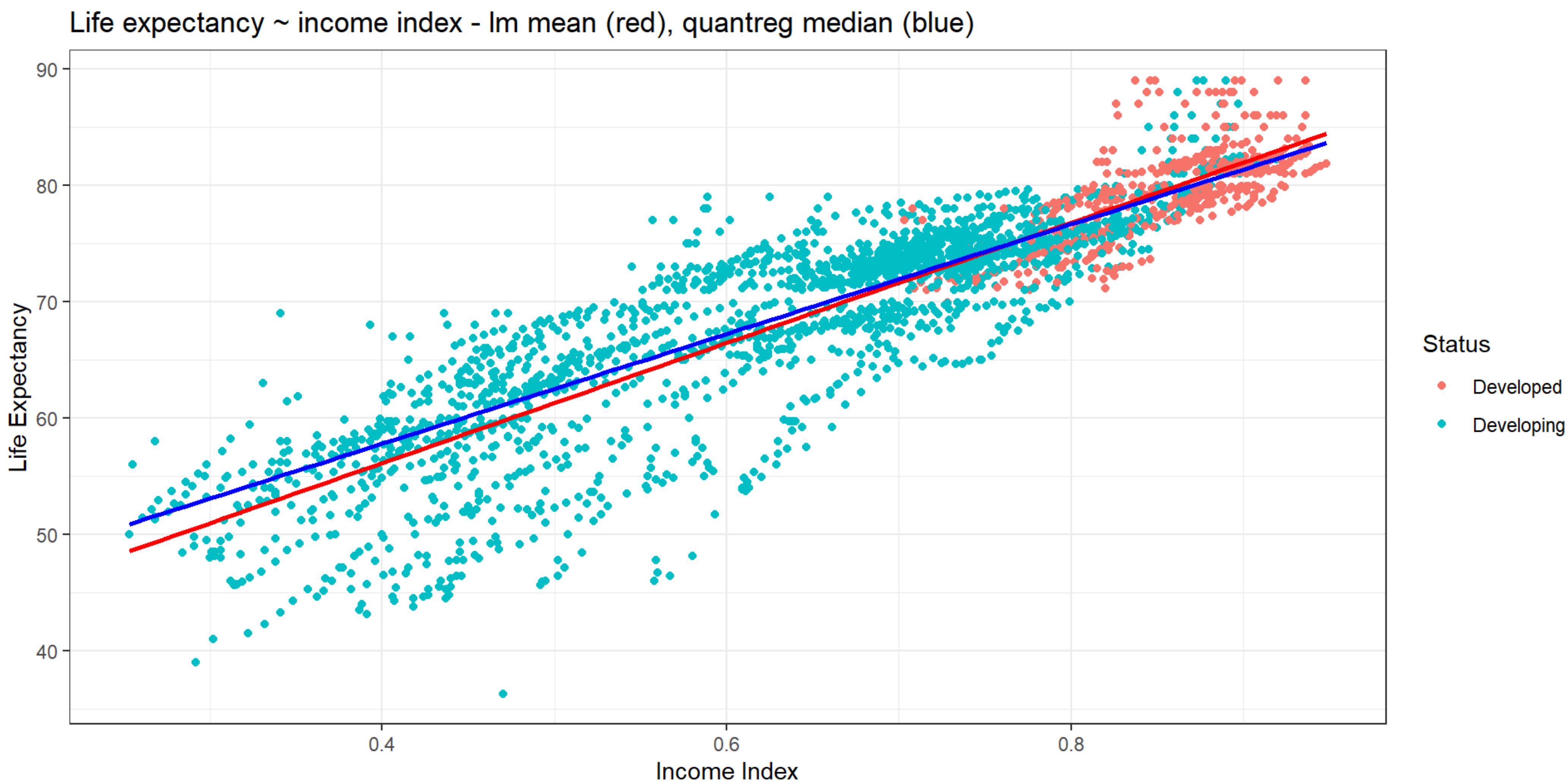
- After removing nations with incomplete information, we can see a positive relationship between developed countries and higher life expectancy.



Quantile regression vs Linear Regression

- While linear regression can give us the mean value, quantreg can allow us to find any quantile ranging from the 1st percentile to the 99th percentile. We can also easily evaluate the 50th percentile (the median) which can be more resistant to outliers than linear regression’s mean.

Quantile regression vs Linear Regression



Quantreg for quantiles

- In addition to comparing the mean to the median, we can also look at the top and bottom quantiles to compare the confidence intervals for life expectancies. First run the different models

Comparing the mean and median

- When comparing the confidence intervals using the median quantile regression and linear regression for the mean, the values roughly align.

```
 Income_composition_of_resources Life_expectancy    Status
 Min.   :0.253                   Min.   :36.30   Length:2349
 1st Qu.:0.525                   1st Qu.:63.70   Class :character
 Median :0.695                   Median :72.50   Mode  :character
 Mean   :0.663                   Mean   :69.69
 3rd Qu.:0.793                   3rd Qu.:76.20
 Max.   :0.948                   Max.   :89.00
```

```
       fit      lwr      upr
1 62.56763 62.37449 62.76078
2 71.34457 71.19324 71.49589
3 76.40421 76.21559 76.59284
```

```
       fit    lower   higher
1 63.67425 63.41876 63.92974
2 71.68521 71.53126 71.83915
3 76.30329 76.13795 76.46863
```

1st Quantile vs 3rd Quantile

- Using the same income scores from the previous slide, we can see that life expectancies are different depending on the quantile of life expectancy. Comparing the 25th quantile to the 75th quantile and the 99th quantile, it’s easy to see how different the life expectancy for those people would be depending on their income. Additionallly, the wealth of an individual’s country plays a factor in life expectancy, regardless of their personal income.

1st Quantile vs 3rd Quantile Prediction Results

- The first set of results is using the quantile regression at the 25th quantile, the second uses the 75th quantile, and the 3rd uses the 99th quantile.

```
       fit    lower   higher
1 59.95276 59.53391 60.37160
2 69.18898 68.96388 69.41407
3 74.51339 74.39632 74.63045
       fit    lower   higher
1 66.03854 65.77021 66.30686
2 73.83366 73.68295 73.98437
3 78.32732 78.17938 78.47525
       fit    lower   higher
1 72.61832 71.21171 74.02493
2 80.40458 79.68416 81.12500
3 84.89313 84.36079 85.42547
```

Next Steps

- Revise example
- Add more to background of quantile regression

References

https://www.voxco.com/blog/quantile-regression/

https://drkebede.medium.com/quantile-regression-tutorial-in-r-f2eec72c132b

https://www.r-bloggers.com/2019/01/quantile-regression-in-r-2/