

# Problem Set 6

William Miller

March 11, 2025

## 1 Informative Visualizations

For this problem set, I decided to use data I gathered for PS5, as well as some data from [teamrankings.com](https://www.teamrankings.com) on NBA defensive efficiency.

### 1.1 Plot 1

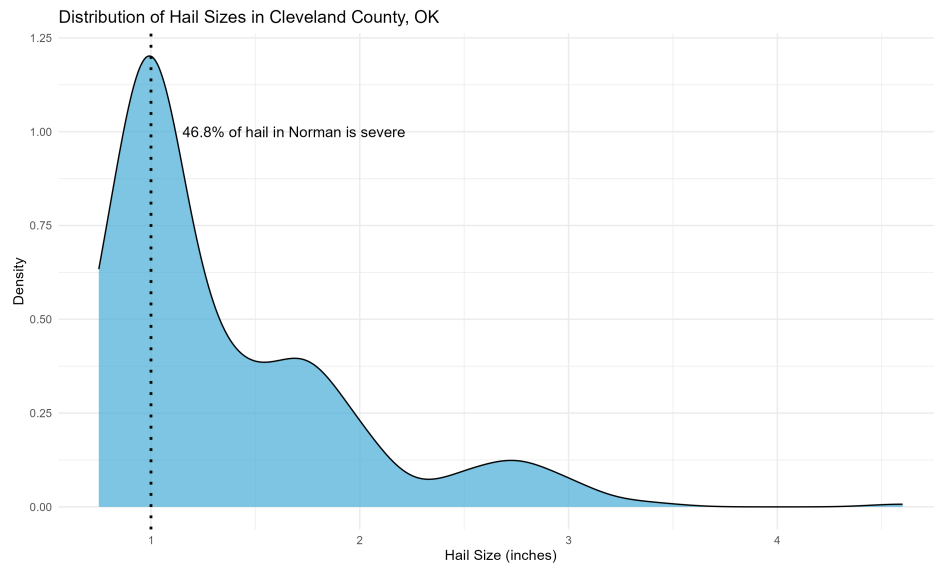


Figure 1: Distribution of Hail Stone Diameter in Cleveland County, OK

This is a plot that shows the distribution of hail stone diameter in Cleveland County, OK. This is data I scraped in problem set 5, and I took steps to clean it there. The main cleaning steps were to use `gsub` in R to remove the units ("in.") from the "Mag" column, and then to convert that column to a numeric variable to be usable in a visualization like this.

To make this specific visualization, I had to calculate the percentage of hail stones with a diameter above 1 inch, which is what the national weather center considers to be a "severe" hail storm. To do that, I calculated the mean of an indicator variable I constructed which took the value 1 if the average hailstone in storm  $i$  was greater than the 1" threshold, and 0 otherwise. I used this measure again in the caption on the plot, referencing it in the text using the `"paste0"` command.

This visualization is useful for understanding hail in Norman because it communicates how severe weather can be in Oklahoma. According to NOAA, hailstones do much more damage as they increase in diameter because they begin to fall much faster as they get larger. Some of the most severe hail storms reported in these data are in the 3+ inch range, and these stones can fall at a speed of up to 72 miles per hour. Understanding the severity of hail storms would be useful in an economic application because homes located in areas that get severe hail regularly would likely be more expensive to insure.

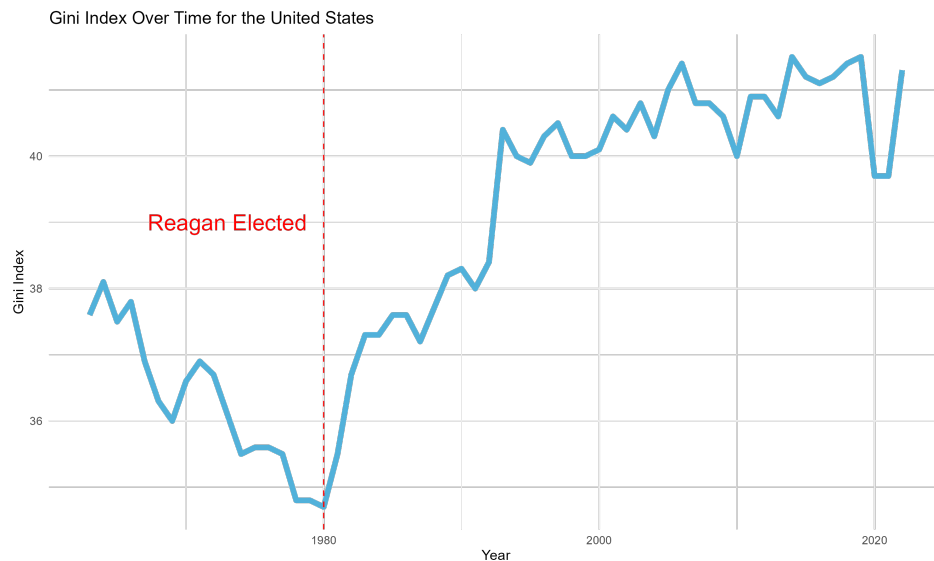


Figure 2: Gini Coefficient in the United States Over Time

## 1.2 Plot 2

This data was also collected in my last problem set. To acquire this data, I used the `fredr` package in R to interface with the FRED API. In the first problem set, I gathered the Gini coefficients for every country in the FRED database. Because they were stored in multiple series, I had to use `fredr` to collect all of the series ids into a list, which I then fed into a custom function that made calls to the FRED API for each series name. Once I had that, I used the `bind_rows` function to put all of the newly collected data into a panel data frame. Finally, I renamed the "value" column to "Gini\_Index" to make it informative and then I removed the prefix "SIPOVGINI" using `gsub` from all of the series names, leaving me with just the iso3 country codes.

For this visualization, I just wanted the USA series, so I used several pipe operators to first filter the "country" column for the series "USA," and then made a `ggplot` based on that filtered dataframe.

This visualization is interesting because it shows how quickly economic inequality rose in the United States after Ronald Reagan was elected. I have heard people say that Reagan's economic policies contributed to the consolidation of economic resources in the United States, and while this simple chart certainly cannot establish causality on its own, it does show that the Reagan's election certainly coincided with the rise of wealth inequality in the US.

## 1.3 Plot 3

This figure shows how defensively dominant the OKC Thunder have been this season. This plot shows that they have the lowest team defensive efficiency rating of any team in the last 5 seasons.

This part of the script is unfortunately not exactly what you were looking for, and you will not be able to execute on your computer. After some reading, it appears the website is built to make scraping challenging. `teamrankings.com` is maintained by BetIQ, which is a company that makes money by selling sports data to bettors. I was unable to properly scrape this data so I collected it manually and put it into an excel spreadsheet before loading that into R.

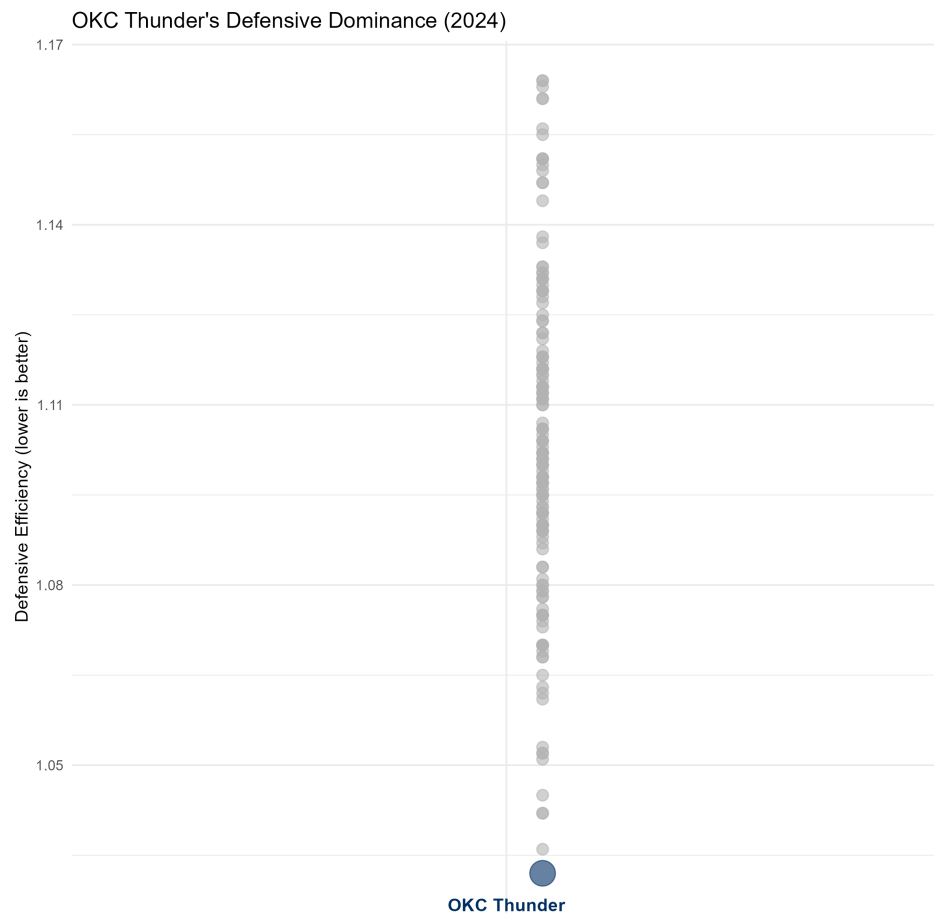


Figure 3: Team Defensive Efficiency Ratings in the NBA in the last 5 seasons