

31_DataVisualization

```
# Load your data files (you can comment out one if not needed)
load("22_processedData.RData")
load("12_cleanedData.RData")

# Convert key columns to correct types

events_data$CPI_Adjusted_Cost_Millions <- as.numeric(events_data$CPI_Adjusted_Cost_Millions)

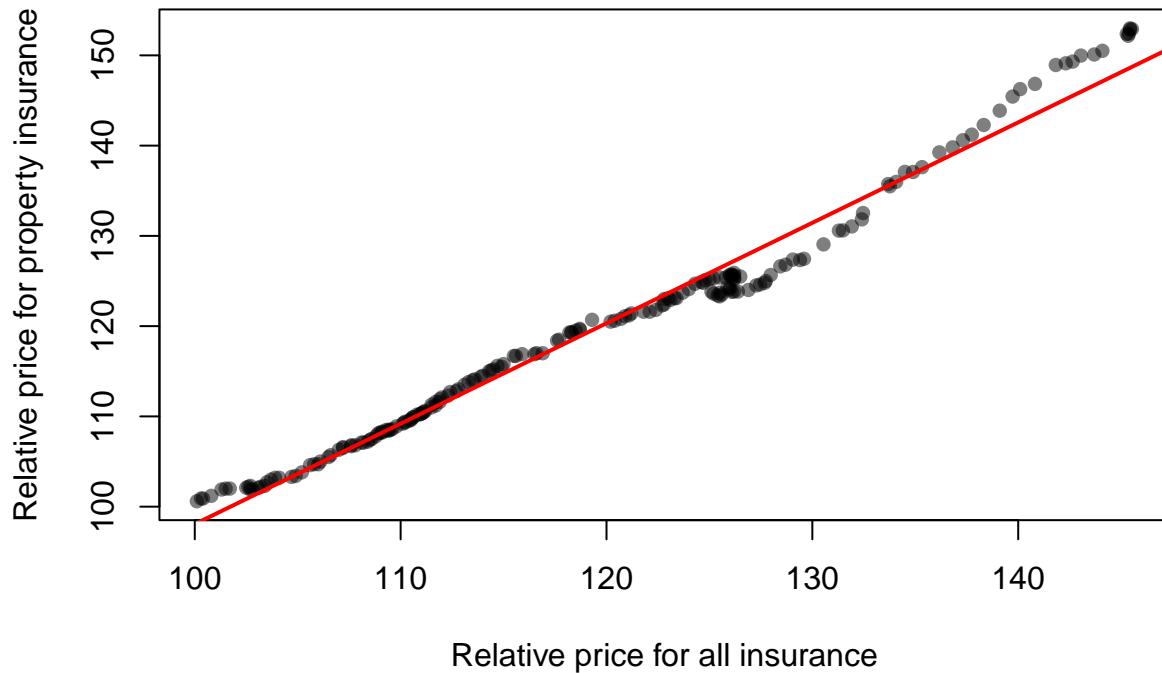
all_insurance$All_Insurance_Index      <- as.numeric(all_insurance$All_Insurance_Index)
all_insurance$Property_Insurance_Index <- as.numeric(all_insurance$Property_Insurance_Index)
```

Visualization

We can compare the property insurance prices to overall insurance prices to see if there's a significant difference between the two.

```
plot(
  all_insurance$All_Insurance_Index,
  all_insurance$Property_Insurance_Index,
  xlab = "Relative price for all insurance",
  ylab = "Relative price for property insurance",
  main = "Property vs All Insurance Indices",
  pch = 16, col = rgb(0,0,0,0.5)
)
abline(lm(Property_Insurance_Index ~ All_Insurance_Index, data = all_insurance),
       col = "red", lwd = 2)
```

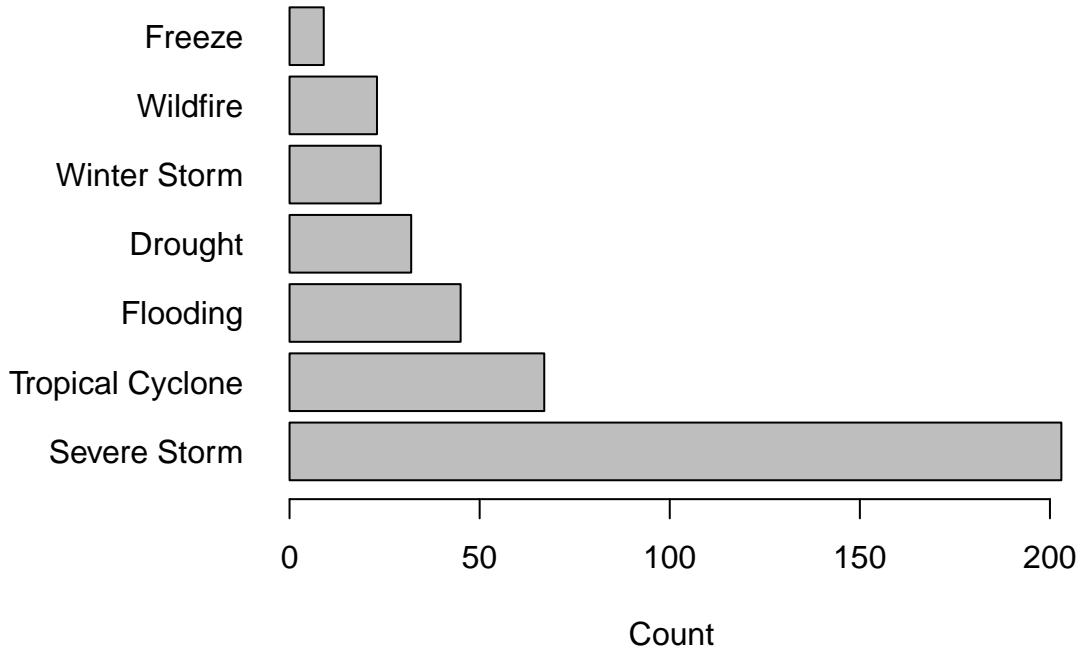
Property vs All Insurance Indices



This scatterplot compares overall insurance prices with property insurance prices. It helps us see whether property insurance behaves differently or just follows the same pattern as the broader market. The points form a clear upward trend, which means both usually rise and fall together. This is expected since damages recorded in our data result in more damage than just property. Because the two move so closely, we might later focus on the difference between them (the “spread”) to find subtle property-specific changes.

```
tbl <- sort(table(events_data$Disaster_Type), decreasing = TRUE)
par(mar = c(5,10,4,2) + 0.1)
barplot(tbl, horiz = TRUE, las = 1,
        main = "Frequency of Major Disasters by Type",
        xlab = "Count")
```

Frequency of Major Disasters by Type



The bar chart shows which natural disaster happens the most within our data set. Seeing that Severe Storms have been recorded the most by a large margin we can no longer generalize that the most total money spent on storms means that they are the most expensive on a one to one ratio. However severe storms have the most overall impact on insurance.

```

## Time Series Analysis

# 1. Extract the year from the 'Insurance_Month' Date column
# We create a new column 'Calc_Year' to aggregate by.
all_insurance$Calc_Year <- format(all_insurance$Insurance_Month, "%Y")

# 2. Annual mean indices from monthly insurance table
# We now aggregate using the newly created 'Calc_Year'
annual_all <- aggregate(all_insurance$All_Insurance_Index ~ all_insurance$Calc_Year,
                           FUN = mean, na.rm = TRUE)
annual_prop <- aggregate(all_insurance$Property_Insurance_Index ~ all_insurance$Calc_Year,
                           FUN = mean, na.rm = TRUE)

# 3. Rename the columns (The grouping column is now the 'Year')
names(annual_all) <- c("Year", "AllIdx")
names(annual_prop) <- c("Year", "PropIdx")

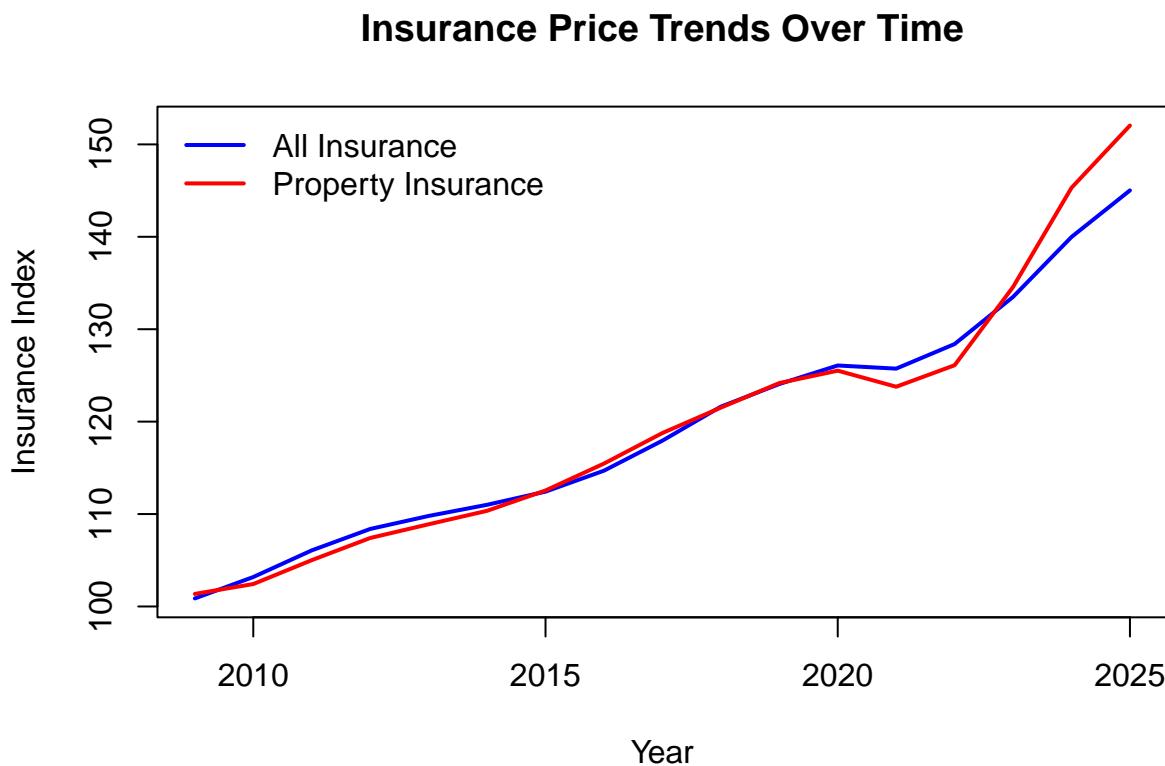
# 4. Align on common years
ii <- merge(annual_all, annual_prop, by="Year")
ii$Year <- as.integer(as.character(ii$Year))
# convert to numeric years

```

```

# 5. Plot the time series
rng <- range(c(ii$AllIdx, ii$PropIdx), na.rm = TRUE)
plot(ii$Year, ii$AllIdx,
  type = "l",
  col = "blue",
  lwd = 2,
  xlab = "Year",
  ylab = "Insurance Index",
  main = "Insurance Price Trends Over Time",
  ylim = rng)
lines(ii$Year, ii$PropIdx,
  col = "red",
  lwd = 2)
legend("topleft",
  legend = c("All Insurance", "Property Insurance"),
  col = c("blue", "red"),
  lwd = 2,
  bty = "n")

```



This line graph tracks how both the overall and property insurance indices change year by year. It's useful for spotting trends and any time periods when property insurance rises faster than the general insurance market. The red (property) line pulls above the blue (overall) line in more recent years indicating that there may be more natural disaster and from what we learned in the last graph more severe storms more recently affecting property.

```

# 1) Histogram of monthly disaster costs
# 2) Boxplot of event cost by type (log-scale)
# 3) Top 6 disaster types by total CPI-adjusted cost
# 4. Align on common years

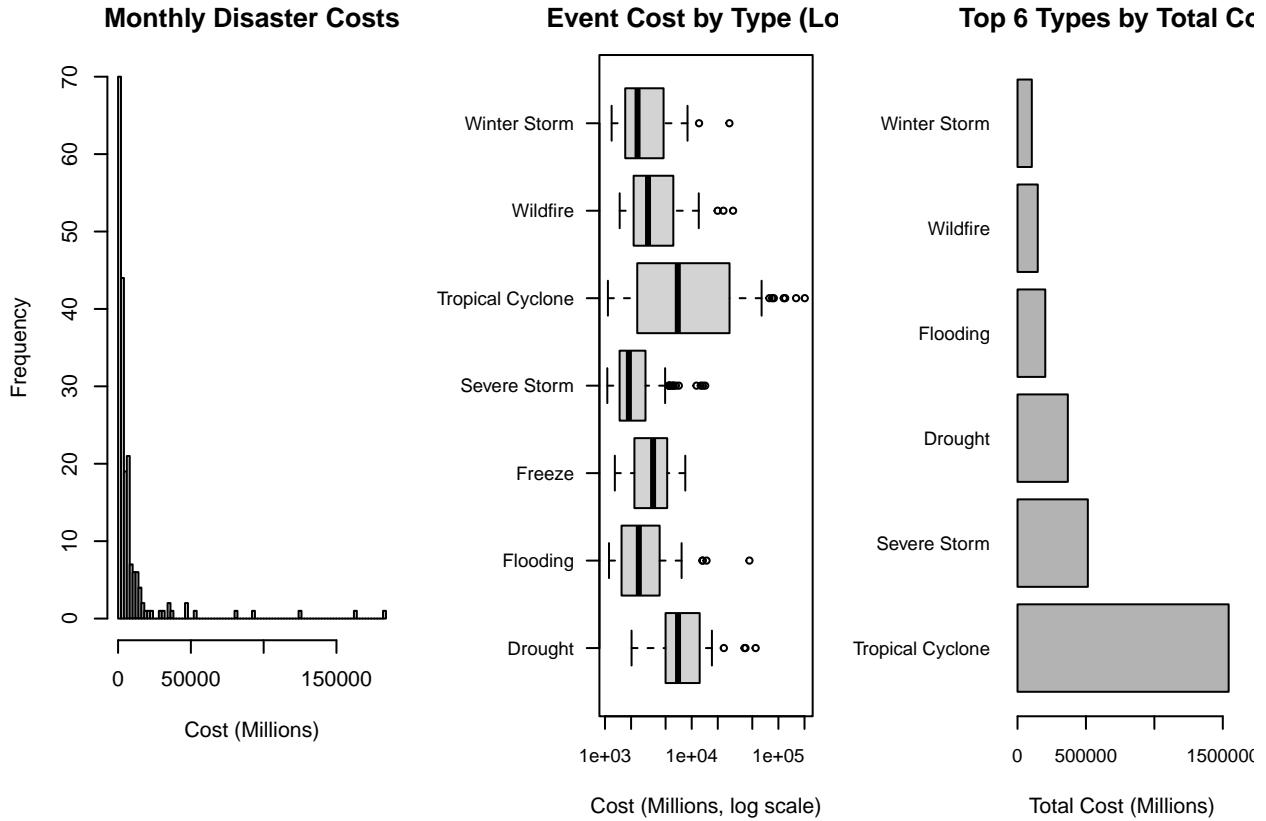
op <- par(no.readonly = TRUE); on.exit(par(op))
par(mfrow = c(1, 3), mar = c(8, 4, 3, 1)) # set layout BEFORE plotting

## (1) Histogram (monthly disaster costs)
hist(all_insurance$Disaster_Cost, breaks = "FD",
      main = "Monthly Disaster Costs",
      xlab = "Cost (Millions)")

## (2) Boxplot by disaster type (horizontal, log-scale)
par(mar = c(5, 7, 3, 1), cex.axis = 0.9) # extra left room for labels
boxplot(CPI_Adjusted_Cost_Millions ~ Disaster_Type, data = events_data,
        horizontal = TRUE, log = "x", las = 1,
        xlab = "Cost (Millions, log scale)",
        ylab = "",
        main = "Event Cost by Type (Log)")

## (3) Top 6 types by total CPI-adjusted cost (rotation fix: horizontal bars)
tot_by_type <- tapply(events_data$CPI_Adjusted_Cost_Millions,
                       events_data$Disaster_Type, sum, na.rm = TRUE)
tot_by_type <- sort(tot_by_type, decreasing = TRUE)
top6 <- head(tot_by_type, 6)
barplot(top6,
        horiz = TRUE, las = 1, xlab = "Total Cost (Millions)",
        main = "Top 6 Types by Total Cost", col = "gray70")

```



Monthly Disaster Costs (Histogram) #1 This histogram shows how disaster costs are distributed each month. Most months have relatively small costs, but a few extreme months stand out with huge spikes. This is predicted since disasters happen at random times. This pattern tells us the data are very right-skewed, meaning a few big events dominate the rest. That's why we use log scales or medians later instead of simple averages, since they can handle big outliers better.

Event Cost by Type (Boxplot) #2 The boxplot compares how expensive different disaster types are. Each box shows the spread and median cost, using a log scale to handle large differences. You can quickly see which types—like hurricanes or severe storms—tend to cause higher damages. This helps decide which disaster categories might have the strongest connection to insurance pricing later on. In a practical sense it tells us what properties not to invest in and insurances the same.

Top 6 Types by Total CPI-Adjusted Cost #3 This chart focuses on the total combined cost of each disaster type. It highlights which ones have the biggest overall financial impact, not just which are common. A few types, such as tropical cyclones and severe storms, make up most of the total losses. These will probably play the biggest role when we study how disasters affect insurance rates since the total costs are so grand. In the grand scheme of disasters there are less overall cyclones so that is important to keep in mind.

```
# Check whether prior disaster costs relate to current insurance index
# Uses 0, 6, and 12 month lags
```

```
op <- par(no.readonly = TRUE); on.exit(par(op))
par(mfrow = c(1, 3), mar = c(4,4,3,1))

ord <- seq_len(nrow(all_insurance))
dc <- all_insurance$Disaster_Cost[ord]
idx <- all_insurance$All_Insurance_Index[ord]
```

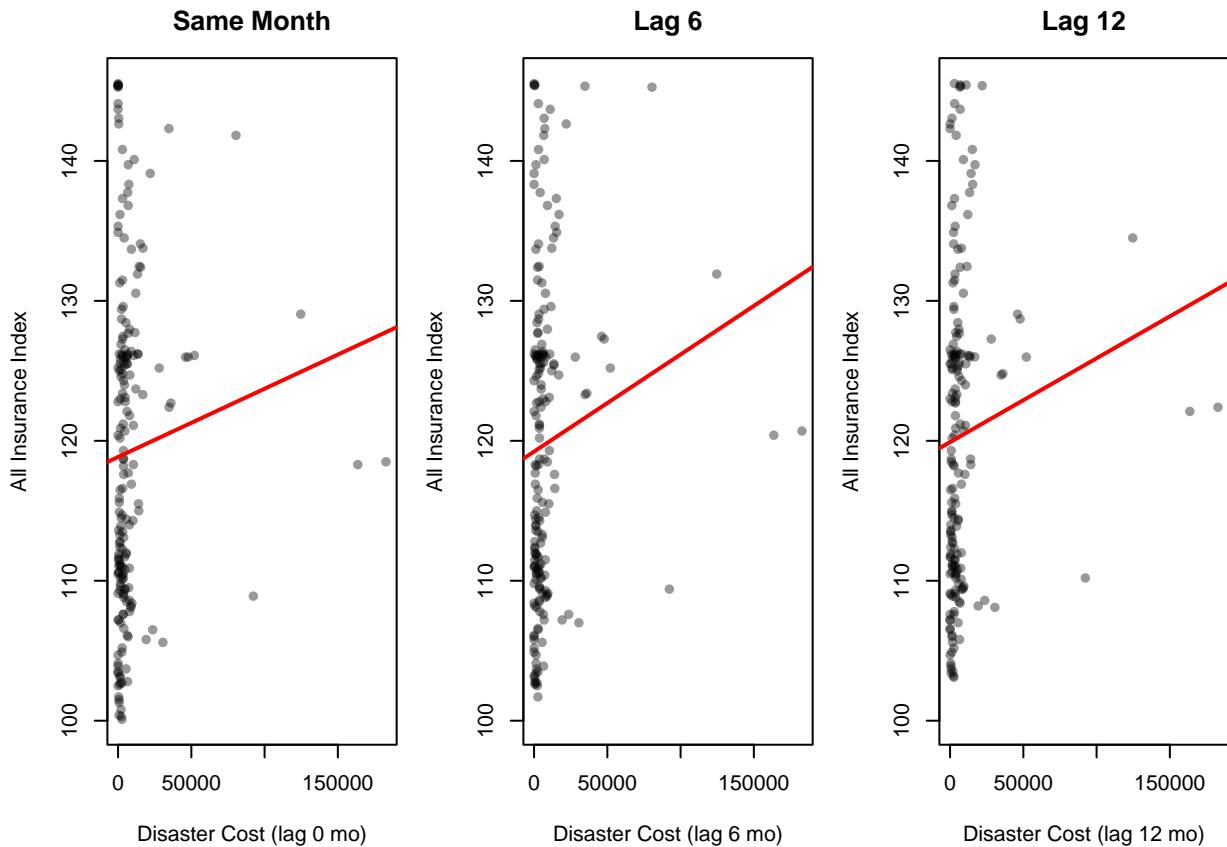
```

lagK <- function(x, k) c(rep(NA, k), head(x, -k))

lags <- list(
  "Same Month" = 0L,
  "Lag 6"       = 6L,
  "Lag 12"      = 12L
)

i <- 1
for (nm in names(lags)) {
  k <- lags[[nm]]
  lx <- if (k == 0L) dc else lagK(dc, k)
  plot(lx, idx,
    xlab = paste0("Disaster Cost (lag ", k, " mo)"),
    ylab = "All Insurance Index",
    main = nm,
    pch = 16, col = rgb(0,0,0,0.4))
  ok <- is.finite(lx) & is.finite(idx)
  if (sum(ok) > 2) abline(lm(idx[ok] ~ lx[ok]), col = "red", lwd = 2)
  i <- i + 1
}

```



Same-Month Disaster Cost vs Insurance Index #1 This scatterplot looks at whether insurance prices go up immediately during months with expensive disasters. The points don't show much of a pattern, which suggests that prices don't react right away. It makes sense—insurers usually adjust rates after assessing the

damage, not in the same month. Also those affected need time to get to safety. It would be unfair to increase insurance rates directly after a severe tragedy.

Lag 6 Months (Disaster Cost vs Insurance Index) #2 Here we shift the disaster costs back by six months to see if there's a delayed effect. The pattern starts to look a bit stronger in correlation, with higher disaster costs linked to slightly higher insurance prices half a year later. This hints that the market might take several months to reflect the impact of big losses. It will likely take even longer to recover from them.

Lag 12 Months (Disaster Cost vs Insurance Index) #3 Finally, we check a one-year lag. The relationship either weakens or stays similar to the six-month one. This helps us understand how long it takes for disasters to influence pricing trends. Whether the response fades after a year or continues. We'll use this to choose the right lag period in our later models. Looking at this graph disasters seem to have a more short term impact on insurance prices overall.