# Case Study

Sarthak Dahal

2024-11-03

```r
library(readxl)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
# Load the Historical Hurricane 1 data
hurricane1 = read_excel("UT Dallas Case Data Clean.xlsx", sheet = "historical_hurricane_1")
head(hurricane1)
```

```
## # A tibble: 6 x 15
##   sid            year number basin subbasin name   iso_time nature   lat   lon
##   <chr>         <dbl>  <dbl> <chr> <chr>    <chr>     <dbl> <chr>  <dbl> <dbl>
## 1 1980161N09249  1980     39 EP    MM       AGATHA   29381  TS      8.9  -111
## 2 1980161N09249  1980     39 EP    MM       AGATHA   29381. TS      8.70 -112.
## 3 1980161N09249  1980     39 EP    MM       AGATHA   29382. TS      8.5  -113.
## 4 1980161N09249  1980     39 EP    MM       AGATHA   29382. TS      8.31 -114.
## 5 1980161N09249  1980     39 EP    MM       AGATHA   29382  TS      8.4  -114.
## 6 1980161N09249  1980     39 EP    MM       AGATHA   29382. TS      8.90 -115.
## # i 5 more variables: wmo_wind_kts <dbl>, wmo_pres_mb <dbl>, wmo_agency <chr>,
## #   track_type <chr>, dist2land_km <dbl>
```

```r
# Load Historical Hurricane 2 data
hurricane2 = read_excel("UT Dallas Case Data Clean.xlsx", sheet = "historical_hurricane_2")
head(hurricane2)
```

```
## # A tibble: 6 x 9
##   storm_name date_time           date                time     longitude latitude
##   <chr>      <dttm>              <dttm>              <chr>        <dbl>    <dbl>
## 1 ALBERTO    1988-08-05 18:00:00 1988-08-05 00:00:00 6:00:00~     -77.5     32
## 2 ALBERTO    1988-08-06 00:00:00 1988-08-06 00:00:00 12:00:0~     -76.2     32.8
## 3 ALBERTO    1988-08-06 06:00:00 1988-08-06 00:00:00 6:00:00~     -75.2     34
```

```
## 4 ALBERTO    1988-08-06 12:00:00 1988-08-06 00:00:00 12:00:0~     -74.6     35.2
## 5 ALBERTO    1988-08-06 18:00:00 1988-08-06 00:00:00 6:00:00~     -73.5     37
## 6 ALBERTO    1988-08-07 00:00:00 1988-08-07 00:00:00 12:00:0~     -72.4     38.7
## # i 3 more variables: quadrant <chr>, wind_speed <dbl>, wind_radius <chr>
```

```
#glimpse(hurricane2)
```

```
# Load Exposures data
exposures = read_excel("UT Dallas Case Data Clean.xlsx", sheet = "exposures")
head(exposures)
```

```
## # A tibble: 6 x 7
##   Location Latitude Longitude `Total Insured Value` Premium
##      <dbl>    <dbl>     <dbl>                 <dbl>   <dbl>
## 1        1       19     -100.               290874.   2644.
## 2        1       19     -100.               296488.   2697.
## 3        1       19     -100.               301944.   2751.
## 4        1       19     -100.               308254.   2806.
## 5        1       19     -100.               313433.   2862.
## 6        1       19     -100.               318636.   2919.
## # i 2 more variables: `Losses - Non Catastrophe` <dbl>, PolicyYear <dbl>
```

```
#glimpse(exposures)
```

```
# Check for missing values in Historical Hurricane 1
sapply(hurricane1, function(x) sum(is.na(x)))
```

```
##          sid         year       number        basin     subbasin         name
##            0            0            0            0            0            0
##     iso_time       nature          lat          lon wmo_wind_kts  wmo_pres_mb
##            0            0            0            0            0         5150
##   wmo_agency   track_type dist2land_km
##            0            0            0
```

```
sapply(hurricane2, function(x) sum(is.na(x)))
```

```
##  storm_name   date_time        date        time   longitude    latitude
##           0           0           0           0           0           0
##    quadrant  wind_speed wind_radius
##           0           0           0
```

```
sapply(exposures, function(x) sum(is.na(x)))
```

```
##                 Location                 Latitude                Longitude
##                        0                        0                        0
##      Total Insured Value                  Premium Losses - Non Catastrophe
##                        0                        0                        0
##               PolicyYear
##                        0
```

# TASKS:

## [Task from Historical Hurricane 1 Sheet]

**This code is for the task: "Create a table summarizing the number of storms by year (1985 to 2020) and storm type."**

What I did: Created a summary table showing the count of storms by year and storm type (1985–2020).

Why I did it: This was a required task from the Historical Hurricane 1 sheet to understand the annual frequency and type of storms over time, which is relevant for assessing trends or patterns in storm occurrences.

How I did it: Filtered the data to include only years from 1985 to 2020.Grouped the data by year and storm type (nature column) and counted the occurrences within each group.

```
# Filter the data to include only years from 1985 to 2020
hurricane1_filtered = hurricane1 %>%
  filter(year >= 1985 & year <= 2020)

# Count the number of storms by year and storm type (nature)
storm_counts = hurricane1_filtered %>%
  group_by(year, nature) %>%
  summarize(count = n(), .groups = 'drop')  # 'nature' is the storm type

# storm_counts
```

**This code is for the task: "Create a table summarizing the maximum wind speed for each storm (1985 to 2020)."**

What I did: Calculated the maximum wind speed for each storm from 1985 to 2020.

Why I did it: This is a required task from the Historical Hurricane 1 sheet, helping us understand the peak intensity for each storm, which can provide insights into storm severity.

How I did it: Filtered data to the relevant years (1985–2020). Grouped by storm ID (sid) and used the max() function to find the highest wind speed for each storm.

```
# Filter the data to include only years from 1985 to 2020
hurricane1_filtered = hurricane1 %>%
  filter(year >= 1985 & year <= 2020)
```

```
# Calculate the maximum wind speed for each storm, including the year
max_wind_speeds = hurricane1_filtered %>%
  group_by(sid, year) %>%  # Include 'year' in grouping if available in hurricane1_filtered
  summarize(max_wind = max(as.numeric(wmo_wind_kts), na.rm = TRUE), .groups = 'drop')
#max_wind_speeds
```

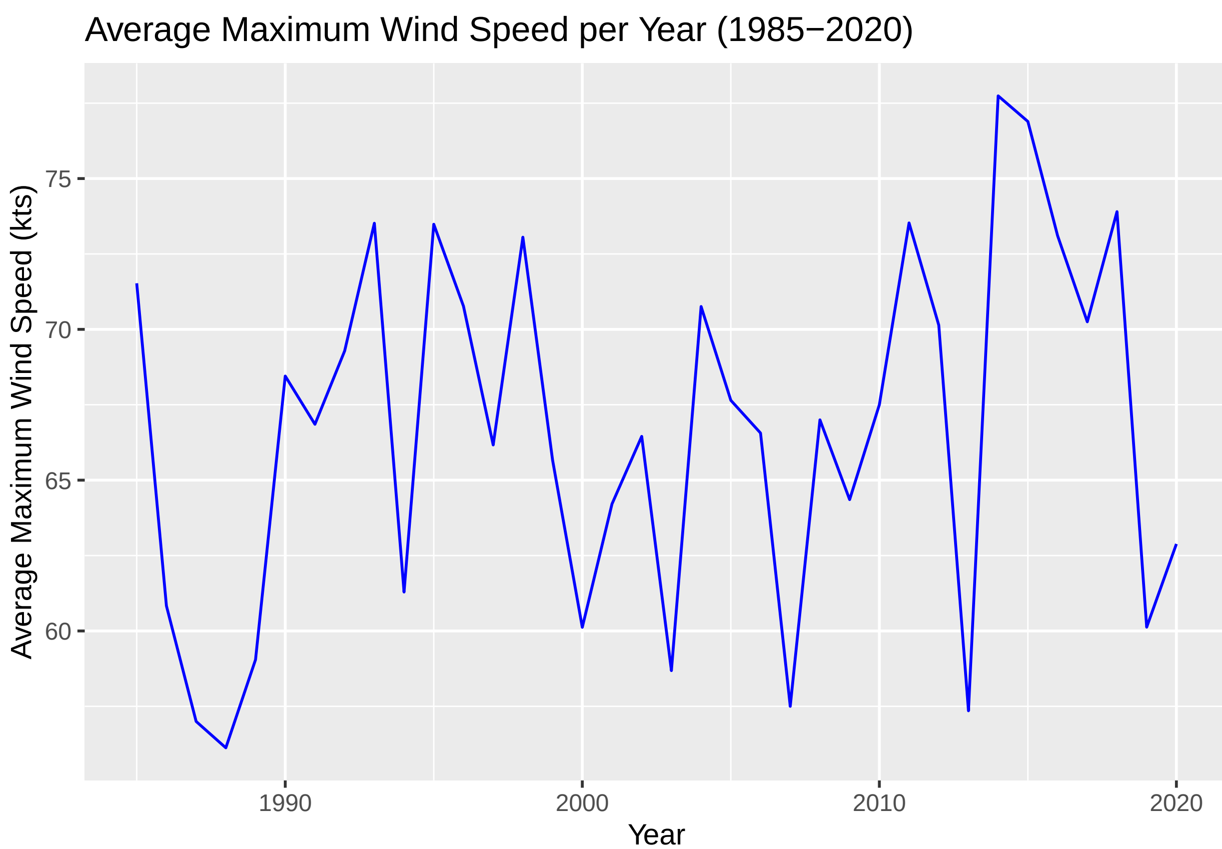
**This code is for the task: "Create a graph that conveys the change in maximum wind speed per storm over time (1985 to 2020)."**

What I did: Calculated the average maximum wind speed per year for all storms and created a line plot.

Why I did it: The original task was to visualize trends in maximum wind speeds over time. Since individual storms only had one observation each, we adjusted to show the trend in average maximum wind speeds by year.

How I did it: Grouped data by year and calculated the average maximum wind speed, then plotted it as a line chart.

```r
# Calculate the average maximum wind speed for all storms in each year
average_max_wind_speed_year = max_wind_speeds %>%
  group_by(year) %>%
  summarize(average_max_wind = mean(max_wind, na.rm = TRUE), .groups = 'drop')

# Plot the average maximum wind speed per year
ggplot(average_max_wind_speed_year, aes(x = year, y = average_max_wind)) +
  geom_line(color = "blue") +
  labs(title = "Average Maximum Wind Speed per Year (1985-2020)", x = "Year", y = "Average Maximum Wind
```




**[Analysis of Findings] ~ [Task from Historical Hurricane 1 Sheet]**

Fluctuations Over Time: The average maximum wind speed per year shows significant fluctuations from 1985 to 2020, with periods of higher and lower intensities.

Notable Peaks: Certain years stand out with particularly high average maximum wind speeds (e.g., around 1990, 2005, and 2010). These peaks could indicate years with stronger storms, which might correlate with increased risk for insured properties in those years.

Potential Trend: Although there's variability, the graph doesn't show a clear increasing or decreasing trend in storm intensity over time. This suggests that storm severity might be influenced by other factors rather than following a simple upward or downward trend over these years.

---

---

# TASKS:

## [Task from Historical Hurricane 2 Sheet]

**This code is for the task: "For each storm, create a new column to pair in Wind (WMO) from Historical Hurricane 1."**

What I did: Paired the maximum wind speed data from Historical Hurricane 1 with each storm in Historical Hurricane 2.

Why I did it: This pairing allows us to compare wind data between the two datasets, checking for consistency in recorded wind speeds for each storm.

How I did it: Calculated the maximum wind speed for each storm in Historical Hurricane 1. Used left_join() to merge this data with Historical Hurricane 2 by matching on storm_name in hurricane2 and sid in max_wind_speeds.

```
#Calculate the maximum wind speed for each storm in Historical Hurricane 1
max_wind_speeds = hurricane1 %>%
  filter(year >= 1985 & year <= 2020) %>%       # Filter to include only relevant years
  group_by(sid) %>%                             # Group by storm ID (sid)
  summarize(max_wind = max(as.numeric(wmo_wind_kts), na.rm = TRUE), .groups = 'drop')  # Calculate max 

# Join the max wind speed data from Historical Hurricane 1 with Historical Hurricane 2
# Using 'storm_name' in hurricane2 and 'sid' in max_wind_speeds to match storms
hurricane2 <- hurricane2 %>%left_join(max_wind_speeds, by = c("storm_name" = "sid"))  # Adds max wind s

#glimpse(hurricane2)
```

## [Task from Historical Hurricane 2 Sheet]

**This code is for the task: "Using the set of storms and the results of the prior steps, create a table that checks the reconciliation of top wind speed between the two sets of data."**

**Important note to understand the output:**

The sid column in Historical Hurricane 1 is a special code that uniquely identifies each storm. - The first four numbers show the year of the storm (e.g., 1985 means the storm happened in 1985). - The rest of the code is just an extra identifier to make each storm unique, especially useful for storms without names. So, sid is like an ID number for each storm to keep track of them, even if we don't know their actual names.

What I did: Created a table that compares the maximum wind speed recorded in Historical Hurricane 1 and Historical Hurricane 2 for each storm.

Why I did it: This comparison allows us to check for consistency in wind speed data across the two datasets, helping us verify the accuracy of the data.

How I did it: Calculated the maximum wind speed for each storm in Historical Hurricane 2. Renamed columns in Historical Hurricane 1 to distinguish between the two datasets. Joined the datasets by storm_name to create a table showing the maximum wind speeds from each source side-by-side.

```r
#Calculate the maximum wind speed for each storm in Historical Hurricane 2
max_wind_speed_hurricane2 = hurricane2 %>%
  group_by(storm_name) %>%
  summarize(max_wind_hurricane2 = max(as.numeric(wind_speed), na.rm = TRUE), .groups = 'drop')

# Merge the maximum wind speeds from both datasets for comparison. We already have max_wind_speeds
# from Historical Hurricane 1 (renaming for clarity)
max_wind_speed_hurricane1 = max_wind_speeds %>%
  rename(storm_name = sid, max_wind_hurricane1 = max_wind)

#Join the two datasets on storm_name to compare wind speeds
wind_speed_comparison = max_wind_speed_hurricane1 %>%
  left_join(max_wind_speed_hurricane2, by = "storm_name")

head(wind_speed_comparison)
```

```
## # A tibble: 6 x 3
##   storm_name    max_wind_hurricane1 max_wind_hurricane2
##   <chr>                       <dbl>               <dbl>
## 1 1985157N16259                  60                  NA
## 2 1985158N10267                 105                  NA
## 3 1985159N09241                  35                  NA
## 4 1985177N10258                 100                  NA
## 5 1985178N10236                  35                  NA
## 6 1985184N15258                  60                  NA
```

# Management Request 1 Tasks Overview

## The tasks for Management Request 1 are focused on understanding changes in the property portfolio over time.

1.1 Calculate and analyze the Loss Ratio over time. 1.2 Analyze changes in Total Insured Value over time. 1.3 Track changes in Premium and Premium per $100 of Total Insured Value over time. 1.4 Analyze the change in Loss Cost over time.

## [Task from Management Request 1]

## 1.1 Calculate and analyze the Loss Ratio over time.

What I did: Calculated the Loss Ratio for each entry in the Exposures sheet and then summarized it by year.

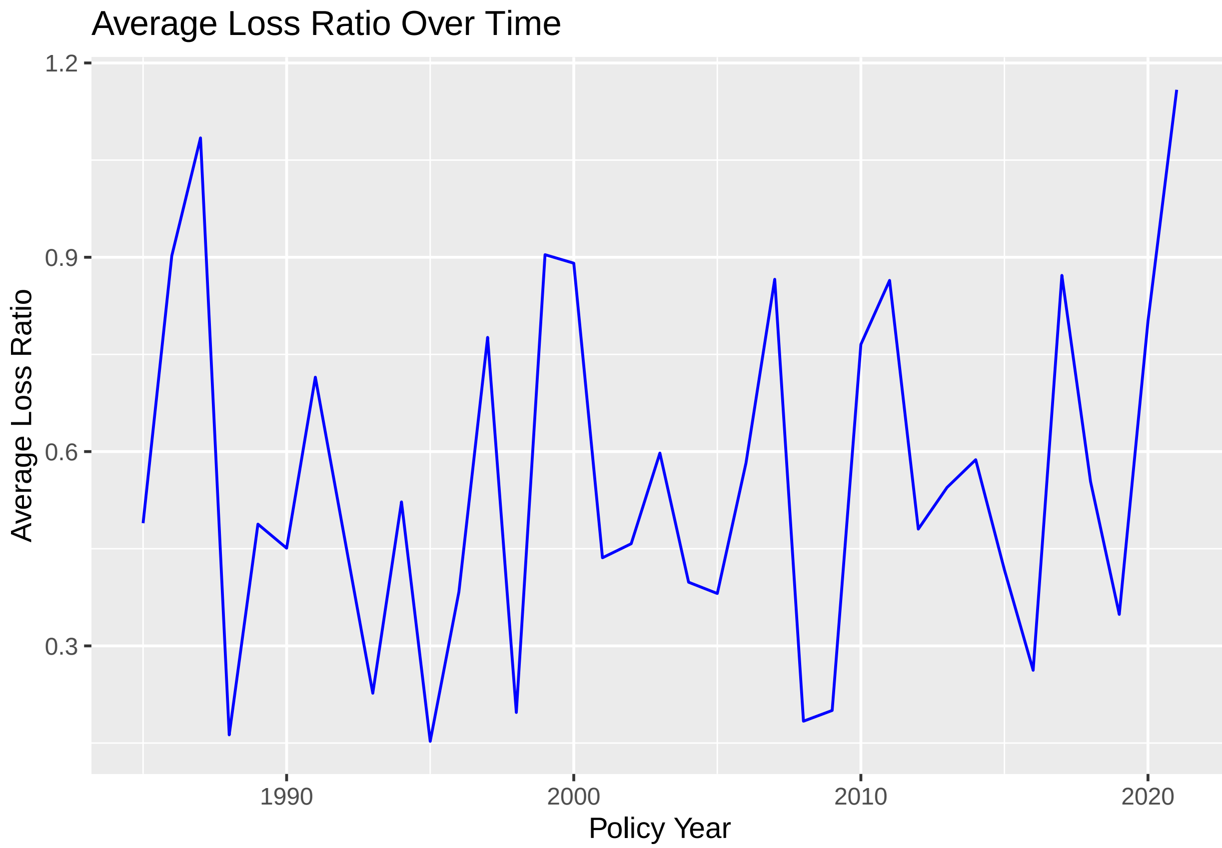
Why I did it: Understanding the Loss Ratio over time shows us how effectively premiums are covering losses, which is important for assessing portfolio performance.

How I did it: Created a new column for Loss Ratio using the formula (Losses/Premium). Grouped by PolicyYear and calculated the average Loss Ratio for each year.

```r
#Calculate Loss Ratio and add it as a new column
exposures = exposures %>%
  mutate(
    loss_ratio = `Losses - Non Catastrophe` / Premium  # Calculate Loss Ratio
  )

#Summarize Loss Ratio over time (by PolicyYear)
loss_ratio_summary = exposures %>%
  group_by(PolicyYear) %>%
  summarize(
    average_loss_ratio = mean(loss_ratio, na.rm = TRUE)  # Average Loss Ratio per year
  )

head(loss_ratio_summary)
```

```
## # A tibble: 6 x 2
##   PolicyYear average_loss_ratio
##        <dbl>              <dbl>
## 1       1985              0.489
## 2       1986              0.902
## 3       1987              1.08
## 4       1988              0.163
## 5       1989              0.488
## 6       1990              0.451
```

```r
ggplot(loss_ratio_summary, aes(x = PolicyYear, y = average_loss_ratio)) +
  geom_line(color = "blue") +
  labs(title = "Average Loss Ratio Over Time",
       x = "Policy Year",
       y = "Average Loss Ratio")
```




**Additional Analysis**

The Average Loss Ratio Over Time plot shows considerable variability from year to year, with values fluctuating between around 0.3 and 1.2. Here's a breakdown of possible insights and approaches:

Initial Observations:

- High Variability: The loss ratio does not appear stable, with some years having much higher loss ratios (e.g., approaching or exceeding 1) and other years having lower ratios.
- Implications of High Ratios: Years with a loss ratio close to or above 1 indicate that losses were nearly equal to or exceeded premiums, which could be a sign of high-risk events or insufficient premium pricing.
- Long-Term Trend: There doesn't seem to be a clear upward or downward trend, suggesting that the loss ratio might be influenced by specific events (e.g., catastrophic losses) rather than showing a general increase or decrease over time.

## Statistical Approach & Analysis

We are setting a hypothetical test to analyze the stability of the loss ratio over time or to see if there's a significant change in recent years compared to earlier years.

Null Hypothesis (H0): There is no significant difference in the average loss ratio between the two periods. Alternative Hypothesis (Ha): There is a significant difference in the average loss ratio between the two periods.

**Summary of why we are doing the T-Test:**

Purpose: To check for a significant shift in the loss ratio over time. Result: The high p-value told us there was no significant difference between the two periods, implying stability in the loss ratio across time.

This approach helps provide evidence that observed fluctuations in the loss ratio are due to random variation, not a clear upward or downward trend in risk.

```
# Split data into two periods
period1 = loss_ratio_summary %>% filter(PolicyYear <= 2000)
period2 = loss_ratio_summary %>% filter(PolicyYear > 2000)

# Perform a two-sample t-test
t_test_result = t.test(period1$average_loss_ratio, period2$average_loss_ratio)
t_test_result
```

```
##
##  Welch Two Sample t-test
##
## data:  period1$average_loss_ratio and period2$average_loss_ratio
## t = -0.096863, df = 29.304, p-value = 0.9235
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1973483  0.1794928
## sample estimates:
## mean of x mean of y
## 0.5510179 0.5599456
```

**Interpretation of the t-test Results**

p-value = 0.9235: This p-value is much higher than a common significance level of 0.05, which means we fail to reject the null hypothesis (H0).

Conclusion: There is no statistically significant difference in the average loss ratio between the two periods (1985–2000 vs. 2001–2020). This suggests that any observed variations in the loss ratio over time are likely due to random fluctuations rather than a consistent upward or downward trend.

# [Task from Management Request 1]

## 1.2 Analyze changes in Total Insured Value over time.

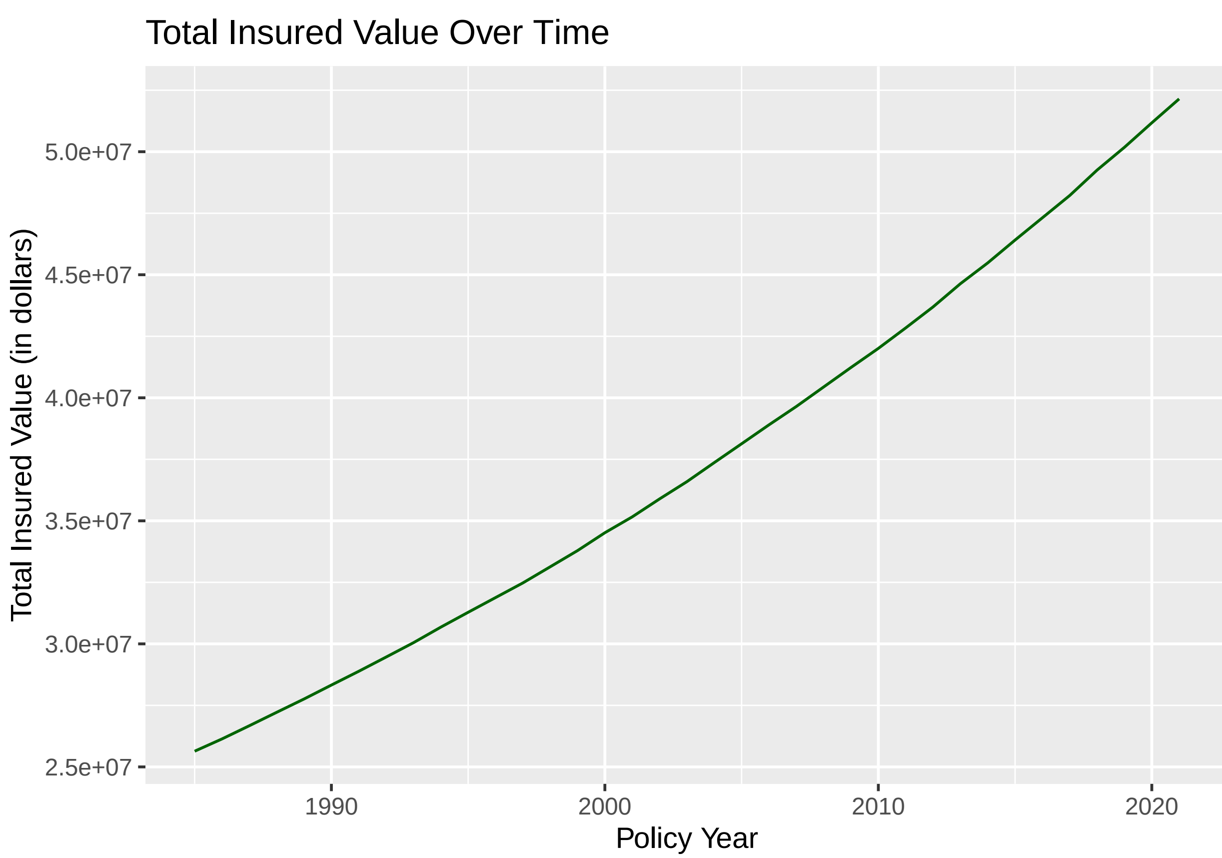
Goal: Track changes in Total Insured Value (TIV) over time to understand how the value of insured properties in the portfolio has evolved.

```
#Summarize Total Insured Value over time (by PolicyYear)
total_insured_value_summary = exposures %>%
  group_by(PolicyYear) %>%
  summarize(
    total_insured_value = sum(`Total Insured Value`, na.rm = TRUE)  # Sum of Total Insured Value per ye
  )

head(total_insured_value_summary)
```

```
## # A tibble: 6 x 2
##   PolicyYear total_insured_value
##        <dbl>               <dbl>
## 1       1985           25637023.
## 2       1986           26136024.
## 3       1987           26672938.
## 4       1988           27218811.
## 5       1989           27759192.
## 6       1990           28325148.
```

**Plotting Total Insured Value Over Time**

```r
# Line plot for Total Insured Value over time
ggplot(total_insured_value_summary, aes(x = PolicyYear, y = total_insured_value)) +
  geom_line(color = "darkgreen") +
  labs(title = "Total Insured Value Over Time",
       x = "Policy Year",
       y = "Total Insured Value (in dollars)")
```




What I did: Calculated the Total Insured Value (sum) for each year to observe the changes over time.

Why I did it: This helps us understand the growth or contraction of the insurance portfolio's value, which is important for assessing the scale of risk coverage over the years.

How I did it: Grouped by PolicyYear and calculated the total of Total Insured Value for each year. Created a line plot to visualize the trend in Total Insured Value over time.

Observations

Consistent Growth: The TIV has steadily increased over the years, indicating that the portfolio of insured properties has grown significantly in terms of value. This could mean that the insurance company has taken on more high-value policies or that the value of insured properties has appreciated over time.

Exponential-like Trend: The upward curve suggests a growth rate that might be exponential, rather than linear. This could be due to factors like rising property values, inflation, or an increase in the number of policies in high-value areas. Implications for Risk Exposure:

As TIV grows, the potential risk exposure also increases because higher TIV means higher potential losses in catastrophic events. This trend could imply a need for adjusting premium pricing, risk management, and reinsurance strategies to account for the increasing portfolio value.

Possible Impact of Economic Factors: The steady increase might reflect broader economic trends, such as property value appreciation or inflation, which would naturally increase the value of insured properties.

## [Task from Management Request 1]

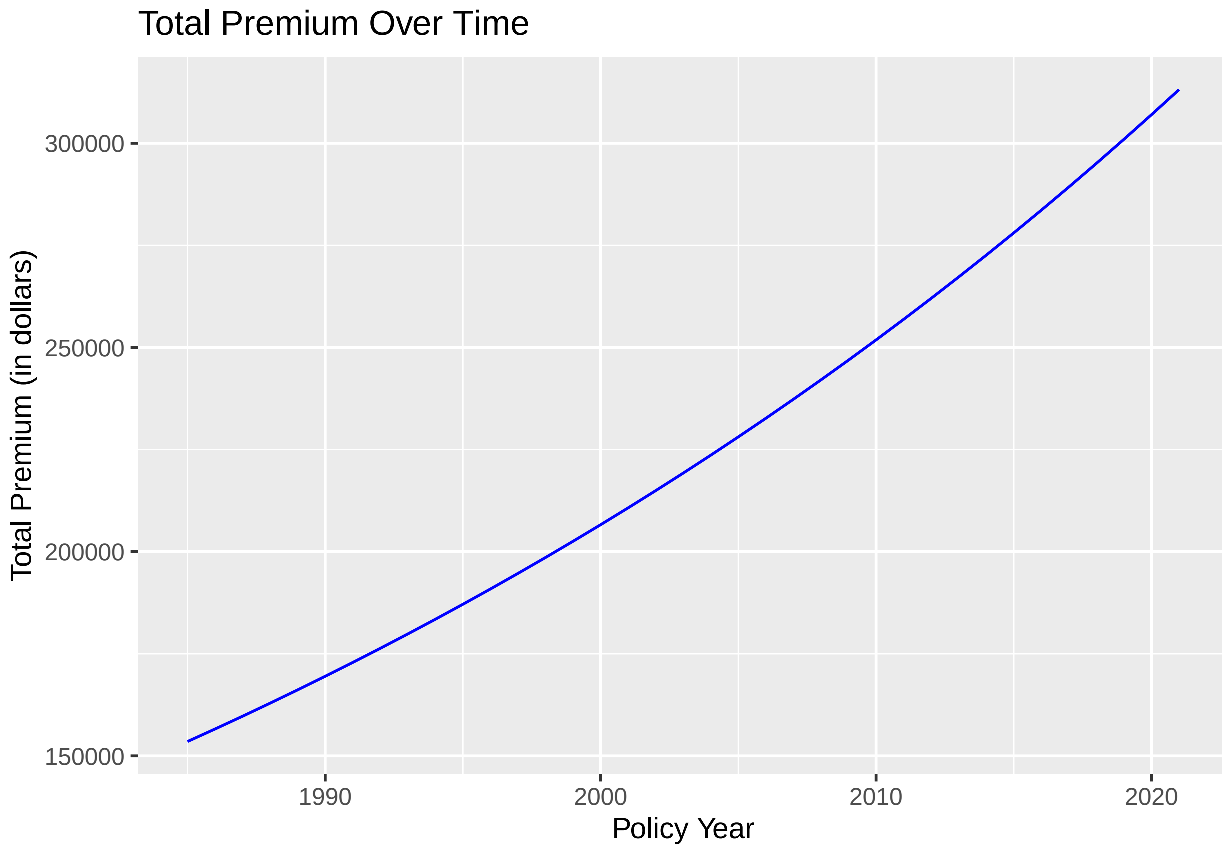
### 1.3 Track changes in Premium and Premium per $100 of Total Insured Value over time.

Goal: Calculate and track Premium per $100 of TIV over time to see if premium growth aligns with the growth in insured value.

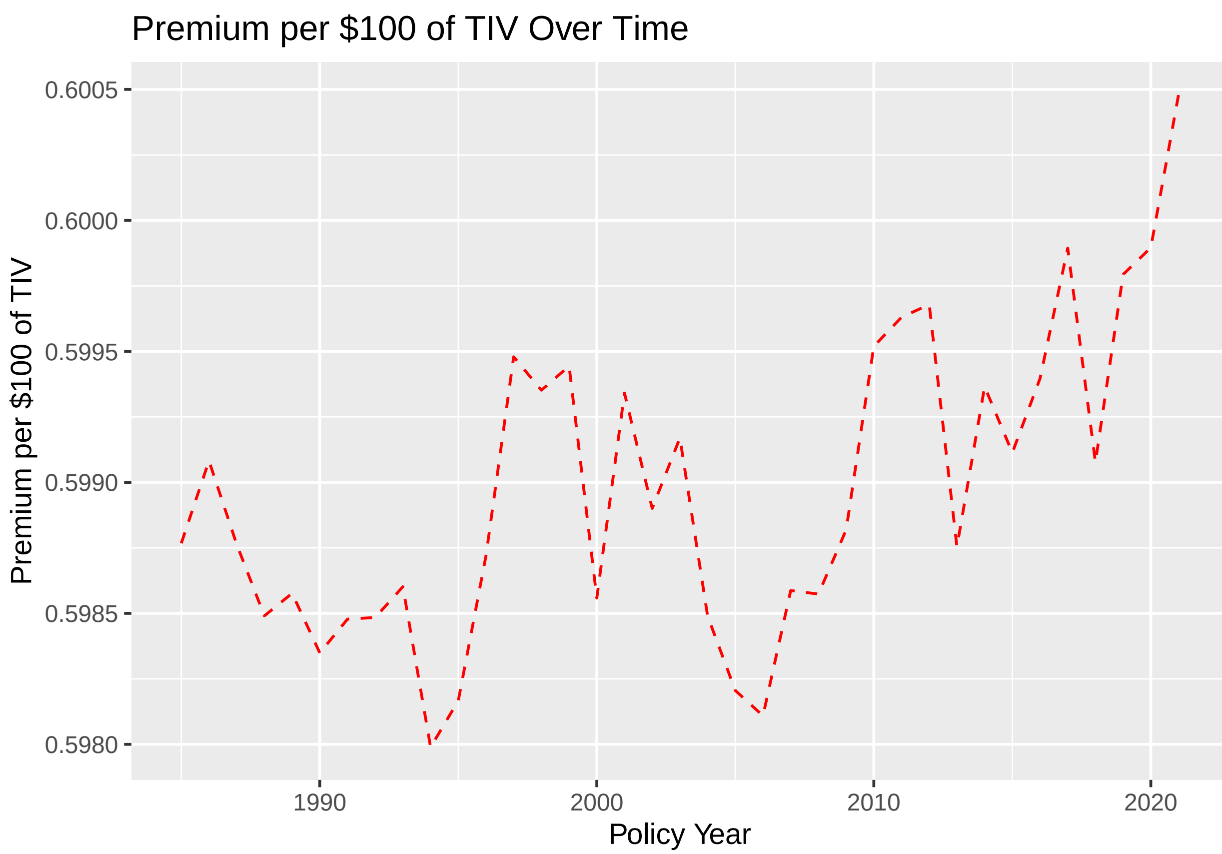
```
#Calculate Premium per $100 of TIV and add it to the summary table
premium_tiv_summary=exposures %>%
  group_by(PolicyYear) %>%
  summarize(
    total_premium = sum(Premium, na.rm = TRUE),                          # Sum of Premium per year
    total_insured_value = sum(`Total Insured Value`, na.rm = TRUE),    # Sum of TIV per year
    premium_per_100_tiv = (sum(Premium, na.rm = TRUE) / sum(`Total Insured Value`, na.rm = TRUE)) * 100
  )

head(premium_tiv_summary)
```

```
## # A tibble: 6 x 4
##   PolicyYear total_premium total_insured_value premium_per_100_tiv
##        <dbl>         <dbl>               <dbl>               <dbl>
## 1       1985       153506.           25637023.               0.599
## 2       1986       156576.           26136024.               0.599
## 3       1987       159708.           26672938.               0.599
## 4       1988       162902.           27218811.               0.598
## 5       1989       166160.           27759192.               0.599
## 6       1990       169483.           28325148.               0.598
```

```
# Plot Total Premium over time
ggplot(premium_tiv_summary, aes(x = PolicyYear, y = total_premium)) +
  geom_line(color = "blue") +
  labs(title = "Total Premium Over Time",
       x = "Policy Year",
       y = "Total Premium (in dollars)")
```




```r
# Plot Premium per $100 of TIV over time
ggplot(premium_tiv_summary, aes(x = PolicyYear, y = premium_per_100_tiv)) +
  geom_line(color = "red", linetype = "dashed") +
  labs(title = "Premium per $100 of TIV Over Time",
       x = "Policy Year",
       y = "Premium per $100 of TIV")
```




Observations

Total Premium Over Time: The Total Premium has been steadily increasing over time, reflecting a growth in the amount collected from policies as the portfolio expands. This trend aligns with the increase in Total Insured Value (TIV), showing that premiums are increasing as more value is insured.

Premium per $100 of TIV: Premium per $100 of TIV shows some fluctuations but is relatively stable overall, with a slight upward trend toward recent years. This suggests that while the total premiums collected are growing, the rate of premium per unit of TIV has been fairly consistent, meaning that the insurance pricing has generally kept up with the growth in insured values.

## [Task from Management Request 1]

### 1.4 Analyze the change in Loss Cost over time.

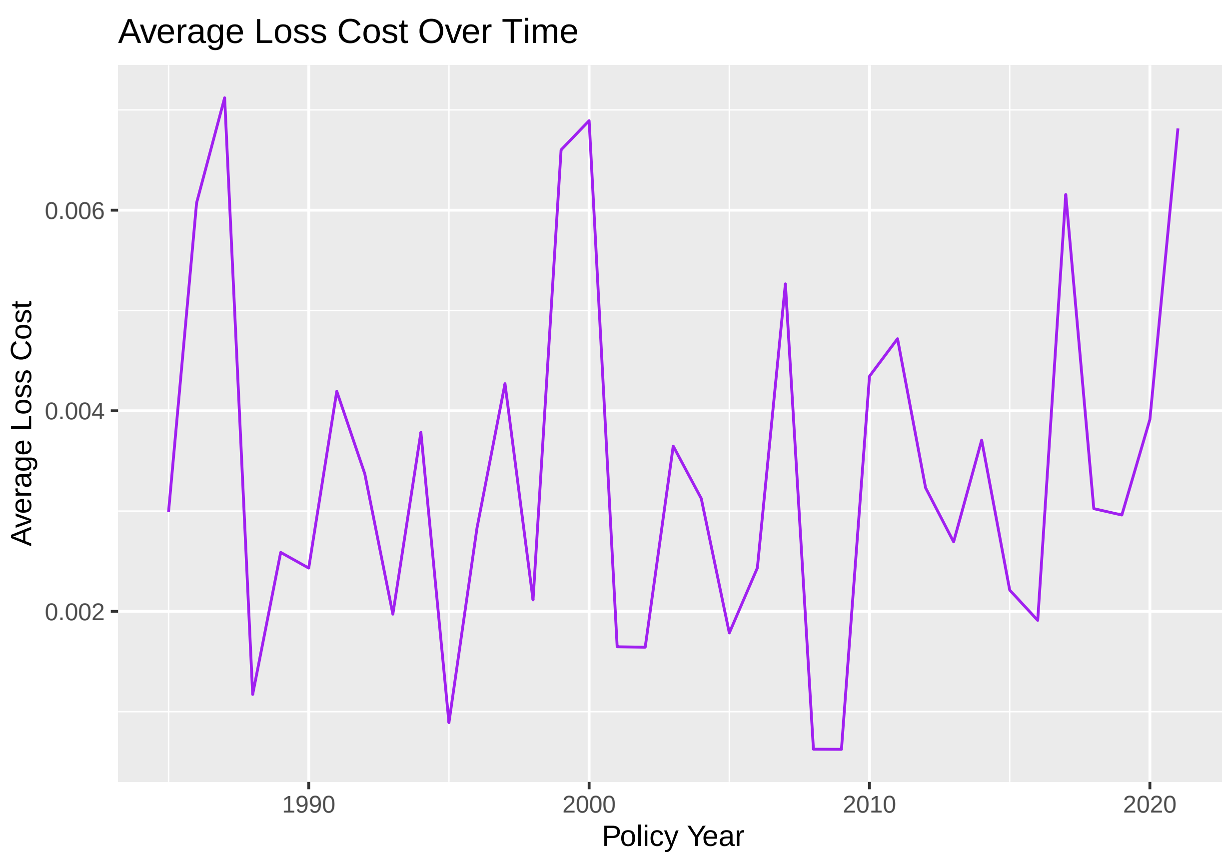
Goal: Calculate and observe Loss Cost over time to understand how losses compare to the Total Insured Value (TIV). This helps assess how much loss occurs per dollar of insured value.

```
#Calculate Loss Cost and add it as a new column
exposures=exposures %>%
  mutate(
    loss_cost = `Losses - Non Catastrophe` / `Total Insured Value`  # Calculate Loss Cost
  )

#Summarize Loss Cost over time (by PolicyYear)
loss_cost_summary=exposures %>%
  group_by(PolicyYear) %>%
```

```r
  summarize(
    average_loss_cost = mean(loss_cost, na.rm = TRUE)  # Average Loss Cost per year
  )

head(loss_cost_summary)
```

```
## # A tibble: 6 x 2
##   PolicyYear average_loss_cost
##        <dbl>             <dbl>
## 1       1985           0.00299
## 2       1986           0.00607
## 3       1987           0.00712
## 4       1988           0.00117
## 5       1989           0.00259
## 6       1990           0.00243
```

```r
# Line plot for Loss Cost over time
ggplot(loss_cost_summary, aes(x = PolicyYear, y = average_loss_cost)) +
  geom_line(color = "purple") +
  labs(title = "Average Loss Cost Over Time",
       x = "Policy Year",
       y = "Average Loss Cost")
```




Observations

Fluctuating Loss Cost: The Loss Cost shows significant fluctuations over time, indicating that the amount of losses per dollar of insured value varies widely from year to year. This variability could be due to differences in claim activity each year, possibly impacted by catastrophic events or other major loss incidents.

No Clear Trend: There doesn't appear to be a clear upward or downward trend in Loss Cost. This suggests that losses relative to TIV have not consistently increased or decreased over time. Instead, they seem to vary around an average level. Higher Peaks in Certain Years:

There are certain years with noticeable peaks in Loss Cost (e.g., the early 1990s, around 2000, and around 2020). These peaks might correspond to years with higher-than-normal losses, potentially due to specific large claims or catastrophic events.

---

---

# Management Request 2 Tasks Overview

## The tasks for Management Request 2 are focused on understanding the risk posed by catastrophes, including the following:

2.1 Concentration of Total Insured Value by geography. 2.2 Current Total Insured Value at risk of hurricane damage (within 1 degree of historic hurricanes). 2.3 Wind speed data analysis for hurricanes impacting the portfolio. 2.4 Categorizing insured locations by probable maximum loss (high, medium, low) for hurricane peril and identifying trends.

## [Task from Management Request 2]

### 2.1 Analyze the change in Loss Cost over time.

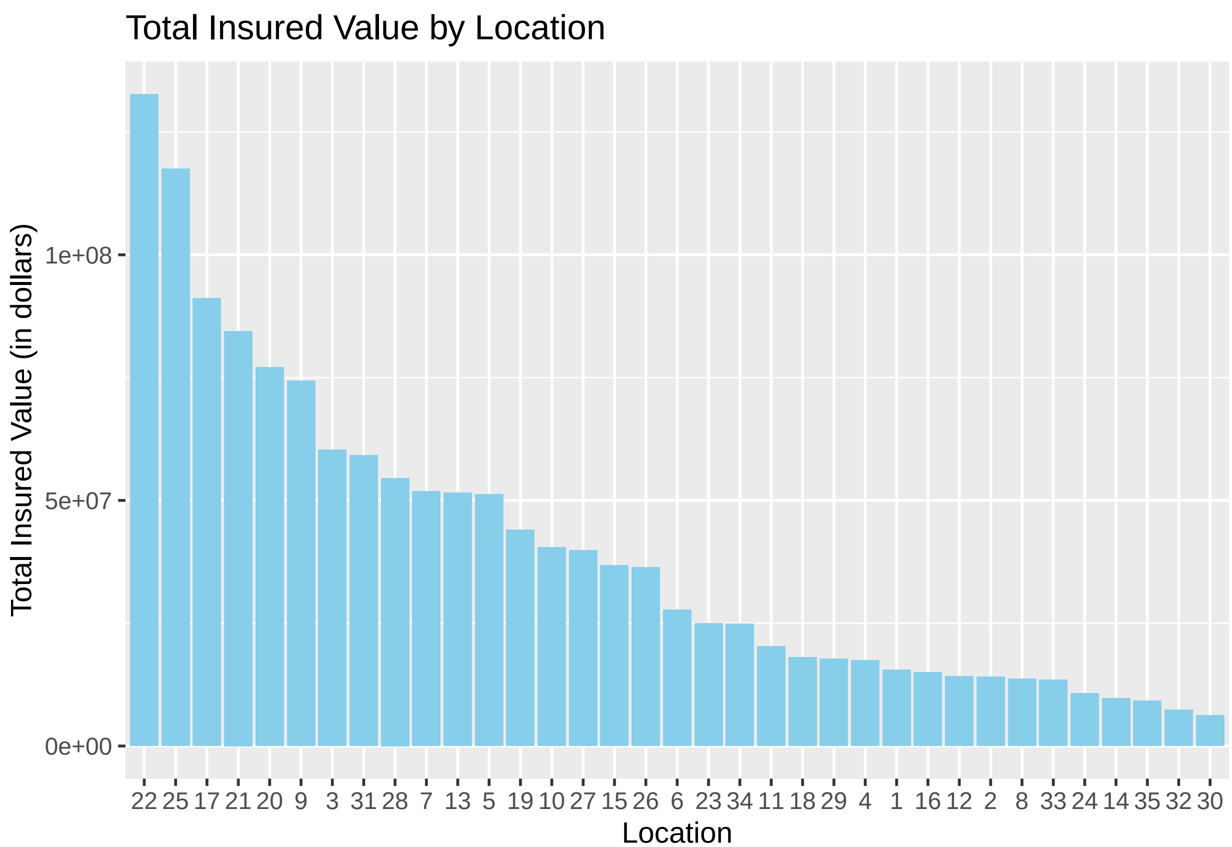
Goal: Determine how TIV is distributed geographically to help identify regions where the insurance company has high exposure.

```r
#colnames(exposures)
#Summarize TIV by Location
tiv_geographic_summary=exposures %>%
  group_by(Location) %>%  # Group by Location
  summarize(
    total_insured_value = sum(`Total Insured Value`, na.rm = TRUE)  # Sum of TIV per location
  )

head(tiv_geographic_summary,10)
```

```
## # A tibble: 10 x 2
##    Location total_insured_value
##       <dbl>               <dbl>
##  1        1           15563192.
##  2        2           14094546.
##  3        3           60339198.
##  4        4           17462758.
##  5        5           51254169.
##  6        6           27737490.
```

```
## 7         7            51862510.
## 8         8            13691106.
## 9         9            74325849.
## 10       10            40469284.
```

```
# Bar plot for TIV concentration by Location
ggplot(tiv_geographic_summary, aes(x = reorder(Location, -total_insured_value), y = total_insured_value
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "Total Insured Value by Location",
       x = "Location",
       y = "Total Insured Value (in dollars)")
```




Observations High Value Concentration: A few locations (like Location 22, 25, and 21) have much higher insured values compared to other locations. This means a large portion of the portfolio's value is concentrated in just a few places.

Risk Implication: If a major event were to impact one of these high-value locations, it could lead to significant losses for the company.

Smaller Locations: Most locations have smaller insured values, contributing less to the total portfolio risk.

## [Task from Management Request 2]

### 2.2 Current Total Insured Value at risk of hurricane damage (within 1 degree of historic hurricanes).

Approach I'll use the data from Historical Hurricane 1 and Historical Hurricane 2 sheets, since they contain latitude and longitude information for each hurricane track. For each property in the exposures dataset, I will check if it is within 1 degree of any hurricane's latitude and longitude. Finally, calculate Total Insured Value at Risk.

```r
#colnames(hurricane1)
#colnames(hurricane2)

# Combine hurricane data from both datasets and select latitude and longitude columns
hurricanes <- bind_rows(
  hurricane1 %>% select(Latitude = lat, Longitude = lon),
  hurricane2 %>% select(Latitude = latitude, Longitude = longitude)
) %>%
  distinct()  # Remove any duplicate coordinates

# Join properties within 1 degree of any hurricane
tiv_at_risk <- exposures %>%
  rowwise() %>%
  mutate(
    at_risk = any(abs(Latitude - hurricanes$Latitude) <= 1 & abs(Longitude - hurricanes$Longitude) <= 1
  ) %>%
  ungroup() %>%
  filter(at_risk) %>%
  summarize(total_insured_value_at_risk = sum(`Total Insured Value`, na.rm = TRUE))

tiv_at_risk
```

```
## # A tibble: 1 x 1
##   total_insured_value_at_risk
##                         <dbl>
## 1                   983174527.
```

**The Total Insured Value at Risk of Hurricane Damage is approximately $98,317,452.**

This value represents the sum of insured values for properties that are located within 1 degree of latitude and longitude of any historic hurricane's track. This amount of TIV is potentially at risk from future hurricanes due to its proximity to previous hurricane paths. The company may want to consider additional risk management strategies for properties within these areas, such as higher premiums, reinsurance, or specific mitigation measures.