Exploring Trends in Global Temperature

Name(English): Hanna

Objective

This project aims to analyze and visualize trends in global temperature over the past century using historical temperature data.

Data Source

You can use the "nycflights13" package in R, which includes a dataset called "weather" containing historical weather data from New York City airports.

Steps

1. Data Exploration: Load the necessary libraries and explore the "weather" dataset to understand its structure and variables. Identify the variables related to temperature and extract the relevant data.

Command

library(tidyverse)

Output

> library(tidyverse)

── **Attaching core tidyverse packages** ────────────────────────────────── tidyverse 2.0.0 ──

✔ dplyr 1.1.4 ✔ readr 2.1.5

✔ forcats 1.0.0 ✔ stringr 1.5.1

✔ ggplot2 3.5.1 ✔ tibble 3.2.1

✔ lubridate 1.9.3 ✔ tidyr 1.3.1

✔ purrr 1.0.2

── **Conflicts** ──────────────────────────────────────────────────── tidyverse\_conflicts() ──

✖ dplyr::filter() masks stats::filter()

✖ dplyr::lag() masks stats::lag()

ℹ Use the conflicted package to force all conflicts to become errors

Explain the command

This command installs and loads the required R packages

Explain the output

Command

library(nycflights13)

weather

Output

# A tibble: 26,115 × 15

origin year month day hour temp dewp humid wind\_dir wind\_speed wind\_gust precip

*<chr>* *<int>* *<int>* *<int>* *<int>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* 1 EWR 2013 1 1 1 39.0 26.1 59.4 270 10.4 NA 0 2 EWR 2013 1 1 2 39.0 27.0 61.6 250 8.06 NA 0 3 EWR 2013 1 1 3 39.0 28.0 64.4 240 11.5 NA 0 4 EWR 2013 1 1 4 39.9 28.0 62.2 250 12.7 NA 0 5 EWR 2013 1 1 5 39.0 28.0 64.4 260 12.7 NA 0 6 EWR 2013 1 1 6 37.9 28.0 67.2 240 11.5 NA 0 7 EWR 2013 1 1 7 39.0 28.0 64.4 240 15.0 NA 0 8 EWR 2013 1 1 8 39.9 28.0 62.2 250 10.4 NA 0 9 EWR 2013 1 1 9 39.9 28.0 62.2 260 15.0 NA 010 EWR 2013 1 1 10 41 28.0 59.6 260 13.8 NA 0# ℹ 26,105 more rows# ℹ 3 more variables: pressure <dbl>, visib <dbl>, time\_hour <dttm># ℹ Use `print(n = ...)` to see more rows

Explain the command

Explain the output

Command

str(weather)

Output

tibble [26,115 × 15] (S3: tbl\_df/tbl/data.frame)

$ origin : chr [1:26115] "EWR" "EWR" "EWR" "EWR" ...

$ year : int [1:26115] 2013 2013 2013 2013 2013 2013 2013 2013 2013 2013 ...

$ month : int [1:26115] 1 1 1 1 1 1 1 1 1 1 ...

$ day : int [1:26115] 1 1 1 1 1 1 1 1 1 1 ...

$ hour : int [1:26115] 1 2 3 4 5 6 7 8 9 10 ...

$ temp : num [1:26115] 39 39 39 39.9 39 ...

$ dewp : num [1:26115] 26.1 27 28 28 28 ...

$ humid : num [1:26115] 59.4 61.6 64.4 62.2 64.4 ...

$ wind\_dir : num [1:26115] 270 250 240 250 260 240 240 250 260 260 ...

$ wind\_speed: num [1:26115] 10.36 8.06 11.51 12.66 12.66 ...

$ wind\_gust : num [1:26115] NA NA NA NA NA NA NA NA NA NA ...

$ precip : num [1:26115] 0 0 0 0 0 0 0 0 0 0 ...

$ pressure : num [1:26115] 1012 1012 1012 1012 1012 ...

$ visib : num [1:26115] 10 10 10 10 10 10 10 10 10 10 ...

$ time\_hour : POSIXct[1:26115], format: "2013-01-01 01:00:00" "2013-01-01 02:00:00" ...

Explain the command

This funtion provides the structure of the dataset weather, includeing the names, types,and first few observations of each variable

Explain the output

2. Data Cleaning: Clean the data by handling missing values, outliers, and any inconsistencies in the dataset that may affect the analysis.

Command

head(weather)

Output

## # A tibble: 6 × 15

## origin year month day hour temp dewp humid wind\_dir wind\_speed wind\_gust

## <chr> <int> <int> <int> <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>

## 1 EWR 2013 1 1 1 39.0 26.1 59.4 270 10.4 NA

## 2 EWR 2013 1 1 2 39.0 27.0 61.6 250 8.06 NA

## 3 EWR 2013 1 1 3 39.0 28.0 64.4 240 11.5 NA

## 4 EWR 2013 1 1 4 39.9 28.0 62.2 250 12.7 NA

## 5 EWR 2013 1 1 5 39.0 28.0 64.4 260 12.7 NA

## 6 EWR 2013 1 1 6 37.9 28.0 67.2 240 11.5 NA

## # ℹ 4 more variables: precip <dbl>, pressure <dbl>, visib <dbl>,

## # time\_hour <dttm>

Explain the command

This function displays the first few rows of the dataset, allowing us to inspect its contents

Explain the output

Command

sum(is.na(weather))

Output

## [1] 23974

Explain the command

Explain the output

Command

weather <- weather %>%

drop\_na()

Output

## origin year month day

## Length:4980 Min. :2013 Min. : 1.000 Min. : 1.00

## Class :character 1st Qu.:2013 1st Qu.: 3.000 1st Qu.: 8.00

## Mode :character Median :2013 Median : 5.000 Median :16.00

## Mean :2013 Mean : 5.802 Mean :15.51

## 3rd Qu.:2013 3rd Qu.: 9.000 3rd Qu.:23.00

## Max. :2013 Max. :12.000 Max. :31.00

## hour temp dewp humid

## Min. : 0.00 Min. : 12.02 Min. :-9.94 Min. : 13.95

## 1st Qu.: 9.00 1st Qu.: 35.06 1st Qu.:15.08 1st Qu.: 37.29

## Median :13.00 Median : 46.94 Median :26.96 Median : 45.92

## Mean :12.62 Mean : 50.60 Mean :30.79 Mean : 48.70

## 3rd Qu.:17.00 3rd Qu.: 64.94 3rd Qu.:46.04 3rd Qu.: 57.04

## Max. :23.00 Max. :100.04 Max. :75.02 Max. :100.00

## wind\_dir wind\_speed wind\_gust precip

## Min. : 10.0 Min. : 4.603 Min. :16.11 Min. :0.000000

## 1st Qu.:230.0 1st Qu.:13.809 1st Qu.:20.71 1st Qu.:0.000000

## Median :280.0 Median :16.111 Median :24.17 Median :0.000000

## Mean :254.7 Mean :16.791 Mean :25.21 Mean :0.002653

## 3rd Qu.:310.0 3rd Qu.:19.563 3rd Qu.:28.77 3rd Qu.:0.000000

## Max. :360.0 Max. :39.127 Max. :66.75 Max. :0.530000

## pressure visib time\_hour

## Min. : 983.8 Min. : 0.120 Min. :2013-01-01 01:00:00.00

## 1st Qu.:1010.7 1st Qu.:10.000 1st Qu.:2013-03-06 10:00:00.00

## Median :1015.2 Median :10.000 Median :2013-05-17 14:00:00.00

## Mean :1015.5 Mean : 9.738 Mean :2013-06-10 08:43:41.93

## 3rd Qu.:1020.4 3rd Qu.:10.000 3rd Qu.:2013-09-22 15:15:00.00

## Max. :1040.4 Max. :10.000 Max. :2013-12-30 18:00:00.00

Explain the command

Explain the output

summary(weather)

1. Data Analysis: Calculate summary statistics for the temperature variables (e.g., mean, median, standard deviation) and visualize the distribution of temperatures over time using line plots.

Command

weather <- weather %>%

filter(temp > -50 & temp < 50)

summary\_stats <- weather %>%

summarise(

mean\_temp = mean(temp),

median\_temp = median(temp),

sd\_temp = sd(temp)

)

print(summary\_stats)Command

Output

## # A tibble: 1 × 3

## mean\_temp median\_temp sd\_temp

## <dbl> <dbl> <dbl>

## 1 35.9 36.0 8.07

Explain the command

Explain the output

Command

ggplot(weather, aes(x = time\_hour, y = temp)) +

geom\_line() +

labs(title = "Temperature Over Time", x = "Time", y = "Temperature (°C)") +

theme\_minimal()

model <- lm(temp ~ time\_hour, data = weather)

summary(model)

Output

##

## Call:

## lm(formula = temp ~ time\_hour, data = weather)

##

## Residuals:

## Min 1Q Median 3Q Max

## -23.2356 -5.3082 0.4491 6.5592 13.7454

##

## Coefficients:

## Estimate Std. Error t value Pr(>|t|)

## (Intercept) -5.822e+01 1.942e+01 -2.998 0.00274 \*\*

## time\_hour 6.878e-08 1.419e-08 4.848 1.32e-06 \*\*\*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## Residual standard error: 8.035 on 2683 degrees of freedom

## Multiple R-squared: 0.008682, Adjusted R-squared: 0.008313

## F-statistic: 23.5 on 1 and 2683 DF, p-value: 1.321e-06

4. Trend Analysis: Use statistical techniques (e.g., linear regression) to analyze trends in global temperature over the years. Create a regression model to predict future temperature trends based on historical data.

Command

ggplot(weather, aes(x = time\_hour, y = temp)) +

geom\_point() +

geom\_smooth(method = "lm", col = "red") +

labs(title = "Temperature Trend Over Time", x = "Time", y = "Temperature (°C)") +

theme\_minimal()

model\_summary <- summary(model)

print(model\_summary)

Output

##

## Call:

## lm(formula = temp ~ time\_hour, data = weather)

##

## Residuals:

## Min 1Q Median 3Q Max

## -23.2356 -5.3082 0.4491 6.5592 13.7454

##

## Coefficients:

## Estimate Std. Error t value Pr(>|t|)

## (Intercept) -5.822e+01 1.942e+01 -2.998 0.00274 \*\*

## time\_hour 6.878e-08 1.419e-08 4.848 1.32e-06 \*\*\*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## Residual standard error: 8.035 on 2683 degrees of freedom

## Multiple R-squared: 0.008682, Adjusted R-squared: 0.008313

## F-statistic: 23.5 on 1 and 2683 DF, p-value: 1.321e-06

Explain the command

Explain the output

5. Insights and Interpretation: Provide insights and interpretations based on the data analysis and visualizations. Discuss any observed trends, patterns, or anomalies in global temperature trends.