615 HW4

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

1. Downloading the data

```
library(data.table)
file_root <- "https://www.ndbc.noaa.gov/view_text_file.php?filename=44013h"
year <- "2023"
tail <- ".txt.gz&dir=data/historical/stdmet/"
all_years_data <- list()
for (year in 1985:2023) {
  path <- paste0(file_root, year, tail)
  print(paste("Downloading and reading data for year:", year))
  try({
    header <- scan(path, what = 'character', nlines = 1, quiet = TRUE)
    buoy <- fread(path, header = FALSE, skip = 2)
    colnames(buoy) <- header
    buoy[, Year := year]
    all_years_data[[as.character(year)]] <- buoy
  }, silent = TRUE)
}
```

```
## [1] "Downloading and reading data for year: 1985"
## [1] "Downloading and reading data for year: 1986"
## [1] "Downloading and reading data for year: 1987"
## [1] "Downloading and reading data for year: 1988"
## [1] "Downloading and reading data for year: 1989"
## [1] "Downloading and reading data for year: 1990"
## [1] "Downloading and reading data for year: 1991"
## [1] "Downloading and reading data for year: 1992"
## [1] "Downloading and reading data for year: 1993"
## [1] "Downloading and reading data for year: 1994"
## [1] "Downloading and reading data for year: 1995"
## [1] "Downloading and reading data for year: 1996"
## [1] "Downloading and reading data for year: 1997"
## [1] "Downloading and reading data for year: 1998"
## [1] "Downloading and reading data for year: 1999"
## [1] "Downloading and reading data for year: 2000"
```

```
## Warning in fread(path, header = FALSE, skip = 2): Stopped early on line 5114.
## Expected 16 fields but found 17. Consider fill=TRUE and comment.char=. First
## discarded non-empty line: <<2000 08 01 00 78 4.3 5.1 0.58 8.33 5.36 999 1022.9
## 17.3 17.5 15.0 99.0 99.00>>
```

```
## [1] "Downloading and reading data for year: 2001"
## [1] "Downloading and reading data for year: 2002"
```

```
## [1] "Downloading and reading data for year: 2003"
## [1] "Downloading and reading data for year: 2004"
## [1] "Downloading and reading data for year: 2005"
## [1] "Downloading and reading data for year: 2006"
## [1] "Downloading and reading data for year: 2007"
## [1] "Downloading and reading data for year: 2008"
## [1] "Downloading and reading data for year: 2009"
## [1] "Downloading and reading data for year: 2010"
## [1] "Downloading and reading data for year: 2011"
## [1] "Downloading and reading data for year: 2012"
## [1] "Downloading and reading data for year: 2013"
## [1] "Downloading and reading data for year: 2014"
## [1] "Downloading and reading data for year: 2015"
## [1] "Downloading and reading data for year: 2016"
## [1] "Downloading and reading data for year: 2017"
## [1] "Downloading and reading data for year: 2018"
## [1] "Downloading and reading data for year: 2019"
## [1] "Downloading and reading data for year: 2020"
## [1] "Downloading and reading data for year: 2021"
## [1] "Downloading and reading data for year: 2022"
## [1] "Downloading and reading data for year: 2023"
```

```
combined_data <- rbindlist(all_years_data, use.names = TRUE, fill = TRUE)
head(combined_data)
```

```
##       YY    MM    DD    hh    WD  WSPD   GST  WVHT   DPD   APD   MWD    BAR
##    <int> <int> <int> <int> <int> <num> <num> <num> <num> <num> <int>  <num>
## 1:    85     1     1     1    80     4     5    99    99    99   999 1030.0
## 2:    85     1     1     2   100     4     5    99    99    99   999 1030.1
## 3:    85     1     1     3   100     4     5    99    99    99   999 1029.4
## 4:    85     1     1     4   110     4     5    99    99    99   999 1028.6
## 5:    85     1     1     5    90     4     5    99    99    99   999 1027.8
## 6:    85     1     1     6    60     4     6    99    99    99   999 1027.7
##     ATMP  WTMP  DEWP   VIS  Year  YYYY  TIDE    mm   #YY  WDIR  PRES
##    <num> <num> <num> <num> <int> <int> <num> <int> <int> <int> <num>
## 1:   5.1   6.7   999    99  1985    NA    NA    NA    NA    NA    NA
## 2:   5.6   6.6   999    99  1985    NA    NA    NA    NA    NA    NA
## 3:   5.8   6.7   999    99  1985    NA    NA    NA    NA    NA    NA
## 4:   5.8   6.7   999    99  1985    NA    NA    NA    NA    NA    NA
## 5:   5.3   6.7   999    99  1985    NA    NA    NA    NA    NA    NA
## 6:   5.5   6.7   999    99  1985    NA    NA    NA    NA    NA    NA
```

2.Changing the 999 and 99 to NA, since I found some data cannot be 99 so many times.

```
head(combined_data)
```

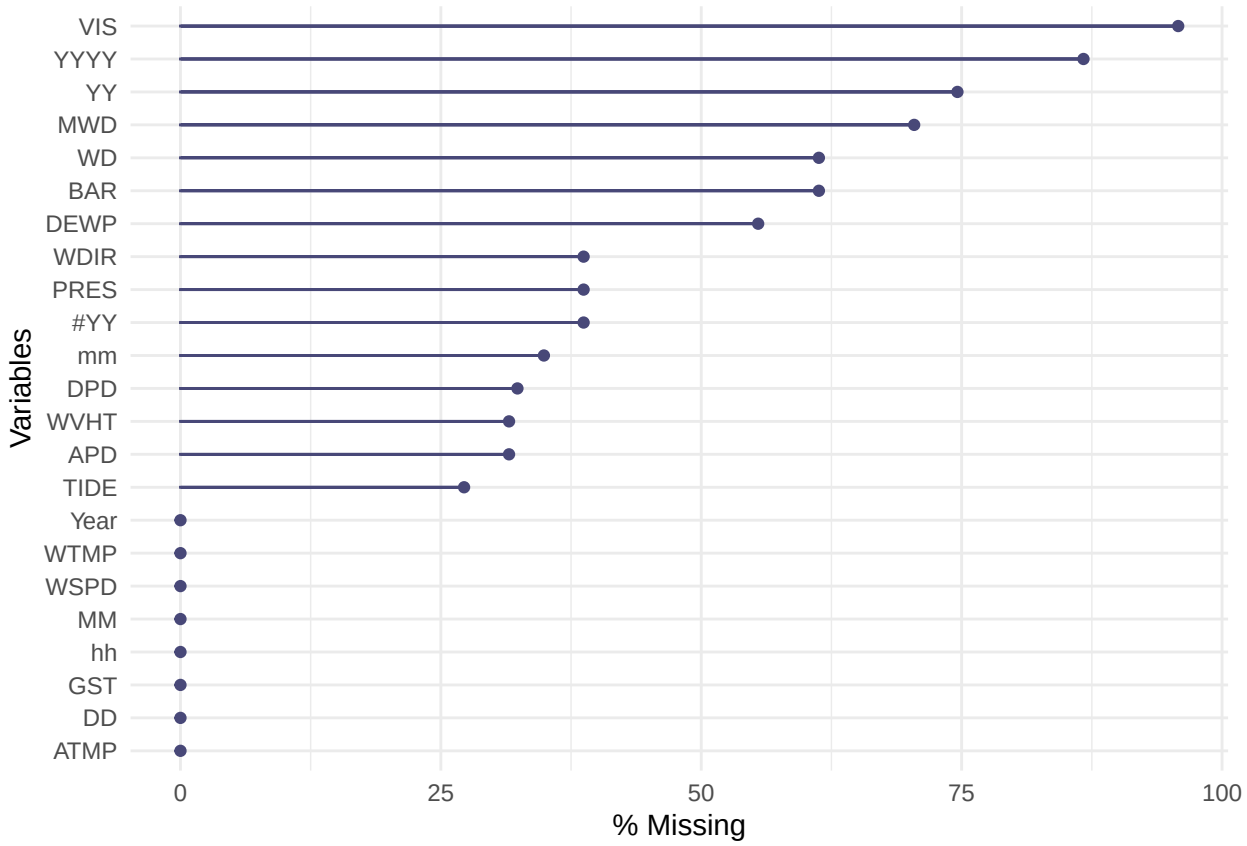
```
##       YY    MM    DD    hh    WD  WSPD   GST  WVHT   DPD   APD   MWD    BAR
##    <int> <int> <int> <int> <int> <num> <num> <num> <num> <num> <int>  <num>
## 1:    85     1     1     1    80     4     5    99    99    99   999 1030.0
## 2:    85     1     1     2   100     4     5    99    99    99   999 1030.1
## 3:    85     1     1     3   100     4     5    99    99    99   999 1029.4
## 4:    85     1     1     4   110     4     5    99    99    99   999 1028.6
## 5:    85     1     1     5    90     4     5    99    99    99   999 1027.8
## 6:    85     1     1     6    60     4     6    99    99    99   999 1027.7
##     ATMP  WTMP  DEWP   VIS  Year  YYYY  TIDE    mm   #YY  WDIR  PRES
##    <num> <num> <num> <num> <int> <int> <num> <int> <int> <int> <num>
```

```
## 1:   5.1   6.7   999    99  1985    NA    NA    NA    NA    NA    NA
## 2:   5.6   6.6   999    99  1985    NA    NA    NA    NA    NA    NA
## 3:   5.8   6.7   999    99  1985    NA    NA    NA    NA    NA    NA
## 4:   5.8   6.7   999    99  1985    NA    NA    NA    NA    NA    NA
## 5:   5.3   6.7   999    99  1985    NA    NA    NA    NA    NA    NA
## 6:   5.5   6.7   999    99  1985    NA    NA    NA    NA    NA    NA
```

```
missing_columns <- c("WVHT", "DPD", "APD", "MWD", "DEWP", "VIS")
for (col in missing_columns) {
  combined_data[[col]][combined_data[[col]] == 999] <- NA
}
for (col in missing_columns) {
  combined_data[[col]][combined_data[[col]] == 99] <- NA
}
head(combined_data)
```

```
##       YY    MM    DD    hh    WD  WSPD   GST  WVHT   DPD   APD   MWD    BAR
##    <int> <int> <int> <int> <int> <num> <num> <num> <num> <num> <int>  <num>
## 1:    85     1     1     1    80     4     5    NA    NA    NA    NA 1030.0
## 2:    85     1     1     2   100     4     5    NA    NA    NA    NA 1030.1
## 3:    85     1     1     3   100     4     5    NA    NA    NA    NA 1029.4
## 4:    85     1     1     4   110     4     5    NA    NA    NA    NA 1028.6
## 5:    85     1     1     5    90     4     5    NA    NA    NA    NA 1027.8
## 6:    85     1     1     6    60     4     6    NA    NA    NA    NA 1027.7
##     ATMP  WTMP  DEWP   VIS  Year  YYYY  TIDE    mm   #YY  WDIR  PRES
##    <num> <num> <num> <num> <int> <int> <num> <int> <int> <int> <num>
## 1:   5.1   6.7    NA    NA  1985    NA    NA    NA    NA    NA    NA
## 2:   5.6   6.6    NA    NA  1985    NA    NA    NA    NA    NA    NA
## 3:   5.8   6.7    NA    NA  1985    NA    NA    NA    NA    NA    NA
## 4:   5.8   6.7    NA    NA  1985    NA    NA    NA    NA    NA    NA
## 5:   5.3   6.7    NA    NA  1985    NA    NA    NA    NA    NA    NA
## 6:   5.5   6.7    NA    NA  1985    NA    NA    NA    NA    NA    NA
```

```
library(naniar)
gg_miss_var(combined_data, show_pct = TRUE)
```



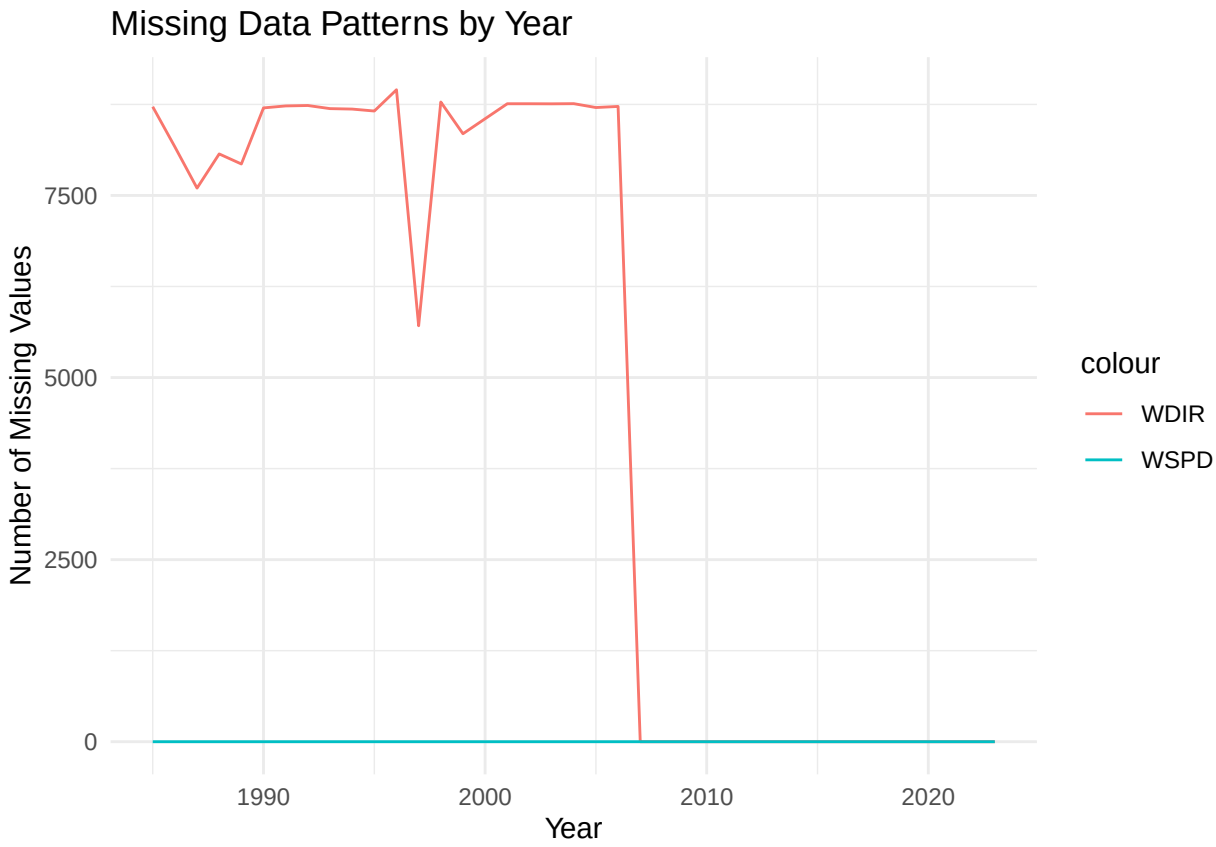

```
missing_by_year <- combined_data[, lapply(.SD, function(x) sum(is.na(x))), by = Year]
print(missing_by_year)
```

```
##       Year    YY    MM    DD    hh    WD  WSPD   GST  WVHT   DPD   APD   MWD
##      <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int>
##  1:  1985     0     0     0     0     0     0     0  8718  8718  8718  8718
##  2:  1986     0     0     0     0     0     0     0  3079  3079  3079  8167
##  3:  1987     0     0     0     0     0     0     0    88    88    88  7601
##  4:  1988     0     0     0     0     0     0     0    53    53    53  8070
##  5:  1989     0     0     0     0     0     0     0   134   135   134  7932
##  6:  1990     0     0     0     0     0     0     0    49    50    49  8702
##  7:  1991     0     0     0     0     0     0     0    15    20    15  8729
##  8:  1992     0     0     0     0     0     0     0    48    48    48  8735
##  9:  1993     0     0     0     0     0     0     0   125   125   125  6676
## 10:  1994     0     0     0     0     0     0     0   141   141   141   281
## 11:  1995     0     0     0     0     0     0     0     8     8     8   158
## 12:  1996     0     0     0     0     0     0     0   328   328   328   474
## 13:  1997     0     0     0     0     0     0     0   344   344   344  1589
## 14:  1998     0     0     0     0     0     0     0   206   206   206  8783
## 15:  1999  8347     0     0     0     0     0     0   113   113   113  8347
## 16:  2001  8759     0     0     0     0     0     0    53    53    53  8759
## 17:  2002  8759     0     0     0     0     0     0    58    58    58  8759
## 18:  2003  8758     0     0     0     0     0     0    90    92    90  8758
## 19:  2004  8760     0     0     0     0     0     0    61    66    61  8760
## 20:  2005  8707     0     0     0     0     0     0    86    86    86  8707
## 21:  2006  8722     0     0     0     0     0     0    63    63    63  8722
## 22:  2007  8694     0     0     0  8694     0     0    64    64    64  8694
```

```
## 23:  2008  8745     0     0     0  8745     0     0    94   298    94  8745
## 24:  2009  8732     0     0     0  8732     0     0    56   237    56  8732
## 25:  2010  7844     0     0     0  7844     0     0    50   227    50  7844
## 26:  2011  8746     0     0     0  8746     0     0   548   715   548  8746
## 27:  2012  4409     0     0     0  4409     0     0   107   233   107   307
## 28:  2013  8209     0     0     0  8209     0     0   356   706   356   817
## 29:  2014  8753     0     0     0  8753     0     0    26   294    26   438
## 30:  2015  8749     0     0     0  8749     0     0    32   385    32   489
## 31:  2016  8695     0     0     0  8695     0     0     8   354     8   454
## 32:  2017  8696     0     0     0  8696     0     0     1   210     1   305
## 33:  2018  8660     0     0     0  8660     0     0     1   245     1   327
## 34:  2019  8613     0     0     0  8613     0     0     0   213     0   293
## 35:  2020 16103     0     0     0 16103     0     0  7559  7820  7559  7926
## 36:  2021 51563     0     0     0 51563     0     0 43279 43595 43279 43691
## 37:  2022 52529     0     0     0 52529     0     0 42835 42939 42835 43074
## 38:  2023 52480     0     0     0 52480     0     0 35344 35503 35344 35729
##      Year    YY    MM    DD    hh    WD  WSPD   GST  WVHT   DPD   APD   MWD
##       BAR  ATMP  WTMP  DEWP   VIS  YYYY  TIDE    mm   #YY  WDIR  PRES
##     <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int>
##  1:     0     0     0  8718  8718  8718  8718  8718  8718  8718  8718
##  2:     0     0     0  8167  8167  8167  8167  8167  8167  8167  8167
##  3:     0     0     0  7601  7601  7601  7601  7601  7601  7601  7601
##  4:     0     0     0  8070  8070  8070  8070  8070  8070  8070  8070
##  5:     0     0     0  7932  7932  7932  7932  7932  7932  7932  7932
##  6:     0     0     0  8702  8702  8702  8702  8702  8702  8702  8702
##  7:     0     0     0  8729  8729  8729  8729  8729  8729  8729  8729
##  8:     0     0     0  8735  8735  8735  8735  8735  8735  8735  8735
##  9:     0     0     0  8692  6651  8692  8692  8692  8692  8692  8692
## 10:     0     0     0  8686   141  8686  8686  8686  8686  8686  8686
## 11:     0     0     0  8659     8  8659  8659  8659  8659  8659  8659
## 12:     0     0     0  8951  8951  8951  8951  8951  8951  8951  8951
## 13:     0     0     0  5711  5711  5711  5711  5711  5711  5711  5711
## 14:     0     0     0  8783  8783  8783  8783  8783  8783  8783  8783
## 15:     0     0     0  3031  8347     0  8347  8347  8347  8347  8347
## 16:     0     0     0    76  8759     0     0  8759  8759  8759  8759
## 17:     0     0     0    48  8759     0     0  8759  8759  8759  8759
## 18:     0     0     0    45  8758     0     0  8758  8758  8758  8758
## 19:     0     0     0    14  8760     0     0  8760  8760  8760  8760
## 20:     0     0     0    39  8707     0     0     0  8707  8707  8707
## 21:     0     0     0  2649  8722     0     0     0  8722  8722  8722
## 22:  8694     0     0  4347  8694  8694     0     0     0     0     0
## 23:  8745     0     0    85  8745  8745     0     0     0     0     0
## 24:  8732     0     0  2727  8732  8732     0     0     0     0     0
## 25:  7844     0     0     4  7844  7844     0     0     0     0     0
## 26:  8746     0     0     9  8746  8746     0     0     0     0     0
## 27:  4409     0     0     4  4409  4409     0     0     0     0     0
## 28:  8209     0     0     7  8209  8209     0     0     0     0     0
## 29:  8753     0     0  5592  8753  8753     0     0     0     0     0
## 30:  8749     0     0  2076  8749  8749     0     0     0     0     0
## 31:  8695     0     0  3440  8695  8695     0     0     0     0     0
## 32:  8696     0     0  8696  8696  8696     0     0     0     0     0
## 33:  8660     0     0  2723  8660  8660     0     0     0     0     0
## 34:  8613     0     0   221  8613  8613     0     0     0     0     0
## 35: 16103     0     0    12 16103 16103     0     0     0     0     0
```

```
## 36: 51563     0     0 27385 51563 51563     0     0     0     0     0
## 37: 52529     0     0 52529 52529 52529     0     0     0     0     0
## 38: 52480     0     0 21669 52480 52480     0     0     0     0     0
##       BAR  ATMP  WTMP  DEWP   VIS  YYYY  TIDE    mm   #YY  WDIR  PRES
```

```
library(ggplot2)
ggplot(missing_by_year, aes(x = Year)) +
  geom_line(aes(y = WDIR, color = "WDIR")) +
  geom_line(aes(y = WSPD, color = "WSPD")) +
  labs(y = "Number of Missing Values", x = "Year", title = "Missing Data Patterns by Year") +
  theme_minimal()
```




```
#Changing 99 or 999 is not always right since there may be some extreme condition which drive the value
#The pattern is that most of the missing value will be detected when time passed, and the most possible
```

3. To showing the climate change, I prefer to use the water temperature(WTMP), the air temperature(ATMP), the wind speed(WSPD), the biometric pressure(BAR), Wave height(WVHT), Average Wave period(APD), Visibility(VIS). I will first calculate the yearly average value of these factors and create a linear rgression line for them to show the transformation clearly. Also, I use a line graph to show the changes.

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

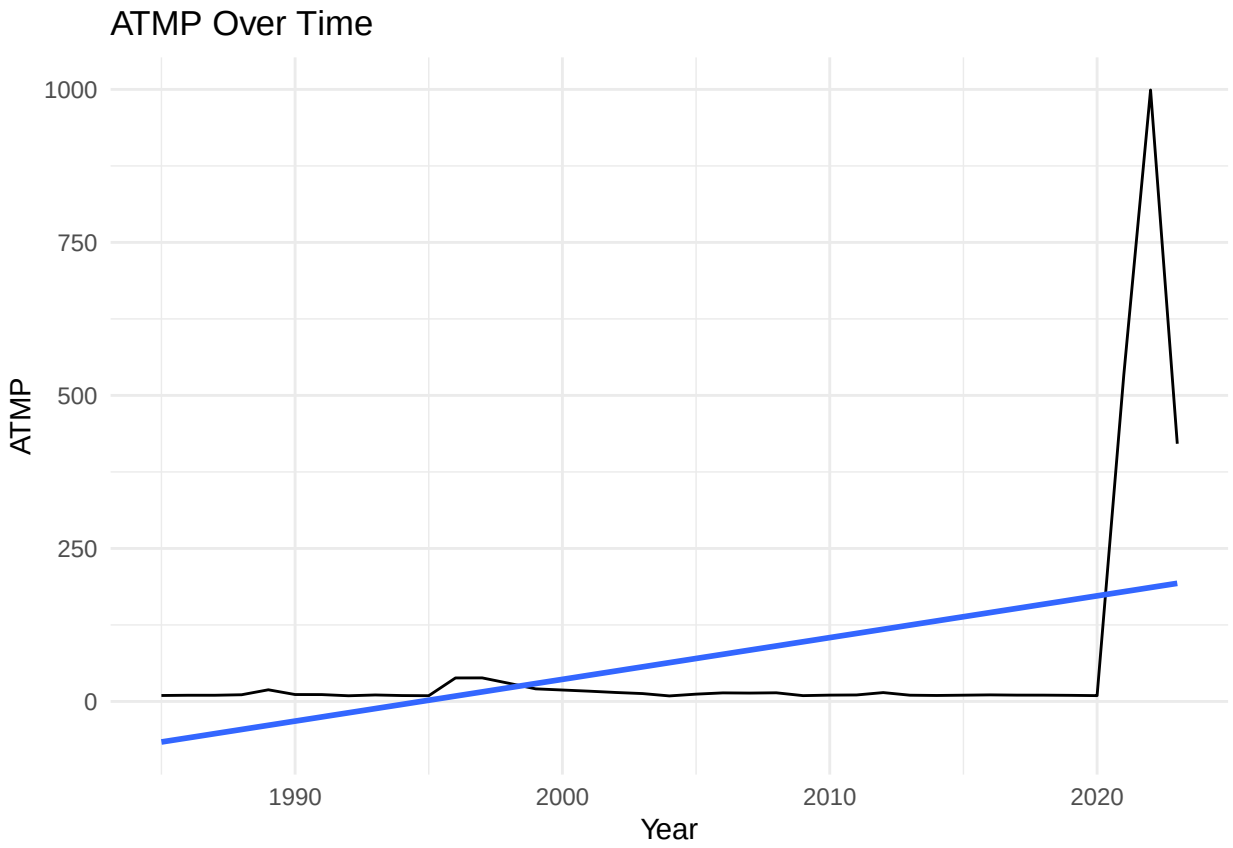
```
## The following objects are masked from 'package:data.table':
##
##     between, first, last
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

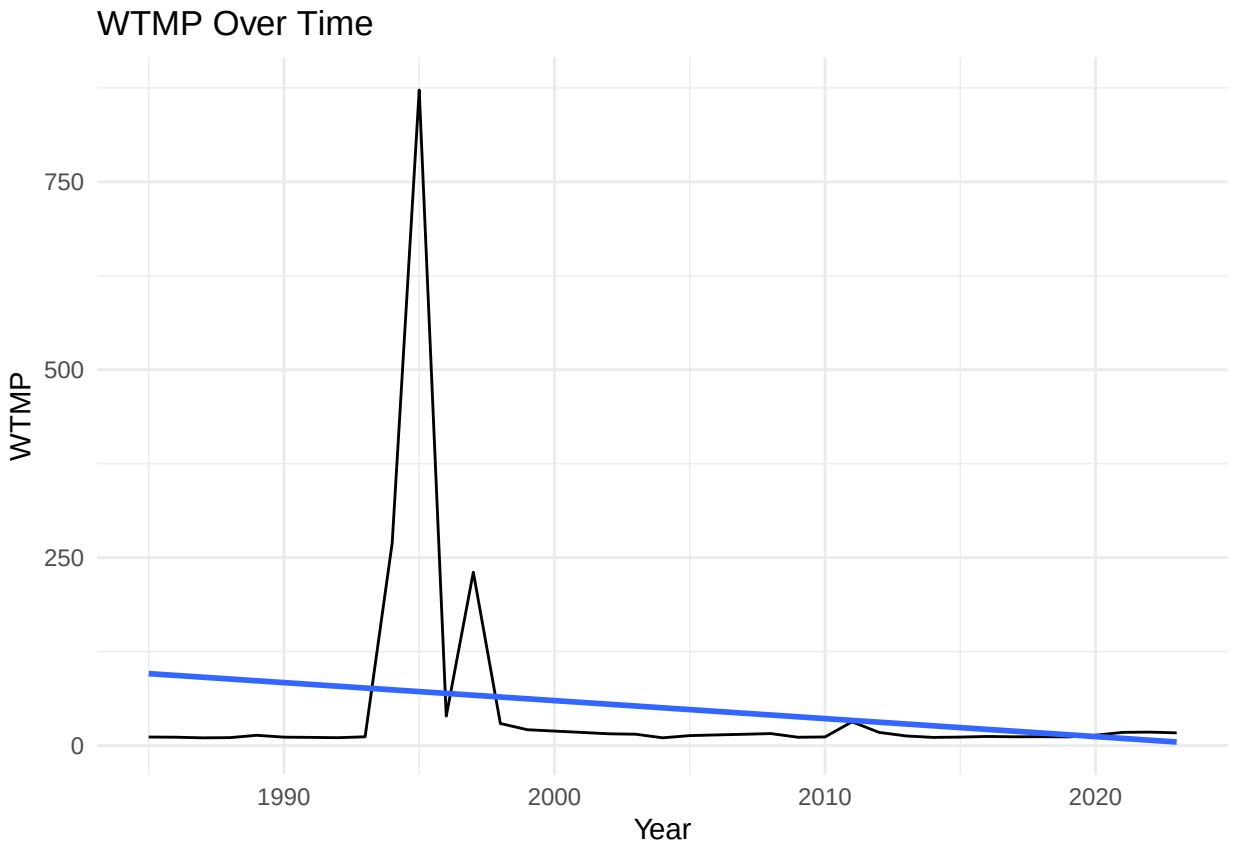
```
yearly_trends <- combined_data %>%
  group_by(Year) %>%
  summarise(
    avg_ATMP = mean(ATMP, na.rm = TRUE),
    avg_WTMP = mean(WTMP, na.rm = TRUE),
    avg_PRES = mean(PRES, na.rm = TRUE),
    avg_WSPD = mean(WSPD, na.rm = TRUE),
    avg_BAR = mean(BAR, na.rm = TRUE),
    avg_VIS = mean(VIS, na.rm = TRUE),
    avg_WVHT = mean(WVHT, na.rm = TRUE),
    avg_APD = mean(APD, na.rm = TRUE)
  )

ggplot(yearly_trends, aes(x = Year, y = avg_ATMP)) +
  geom_line() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "ATMP Over Time",
       x = "Year",
       y = "ATMP") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
ggplot(yearly_trends, aes(x = Year, y = avg_WTMP)) +
  geom_line() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "WTMP Over Time",
       x = "Year",
       y = "WTMP") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

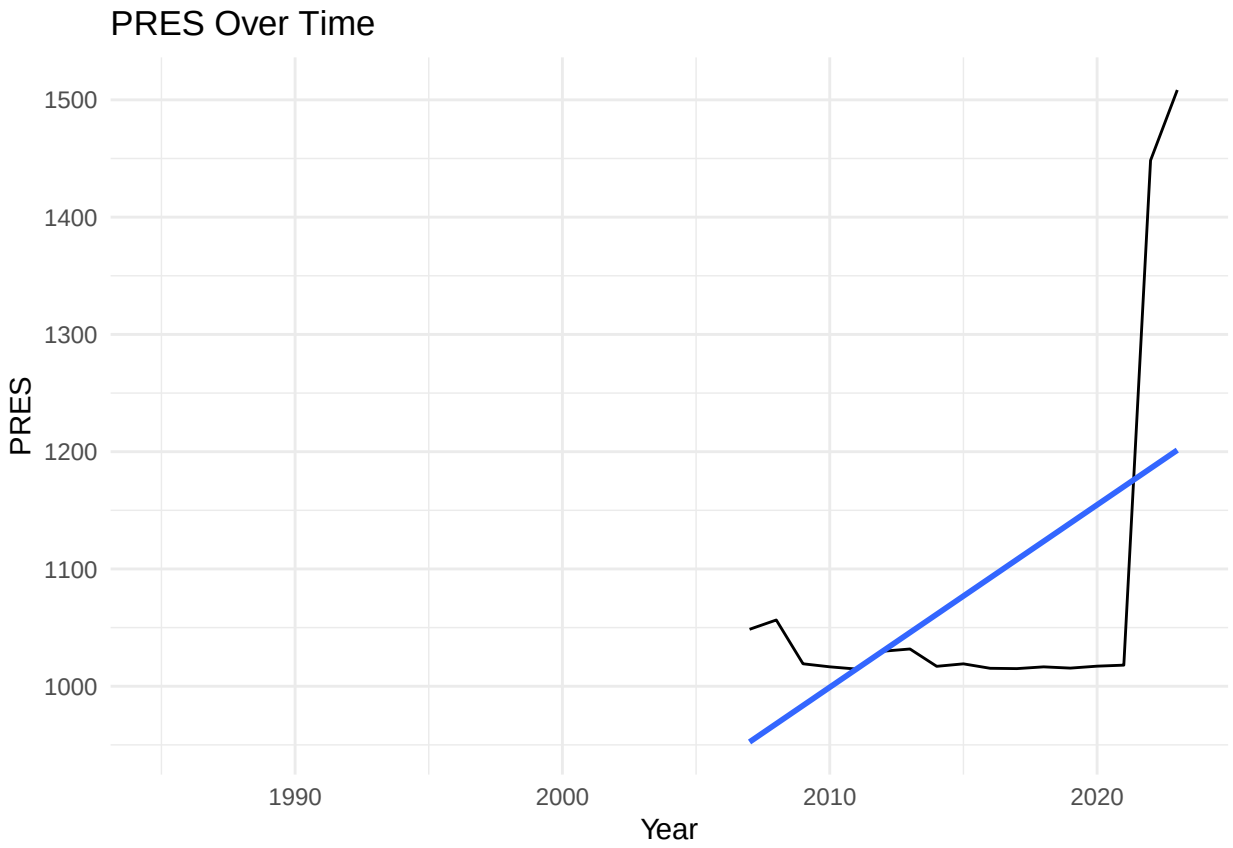



```
ggplot(yearly_trends, aes(x = Year, y = avg_PRES)) +
  geom_line() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "PRES Over Time",
       x = "Year",
       y = "PRES") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

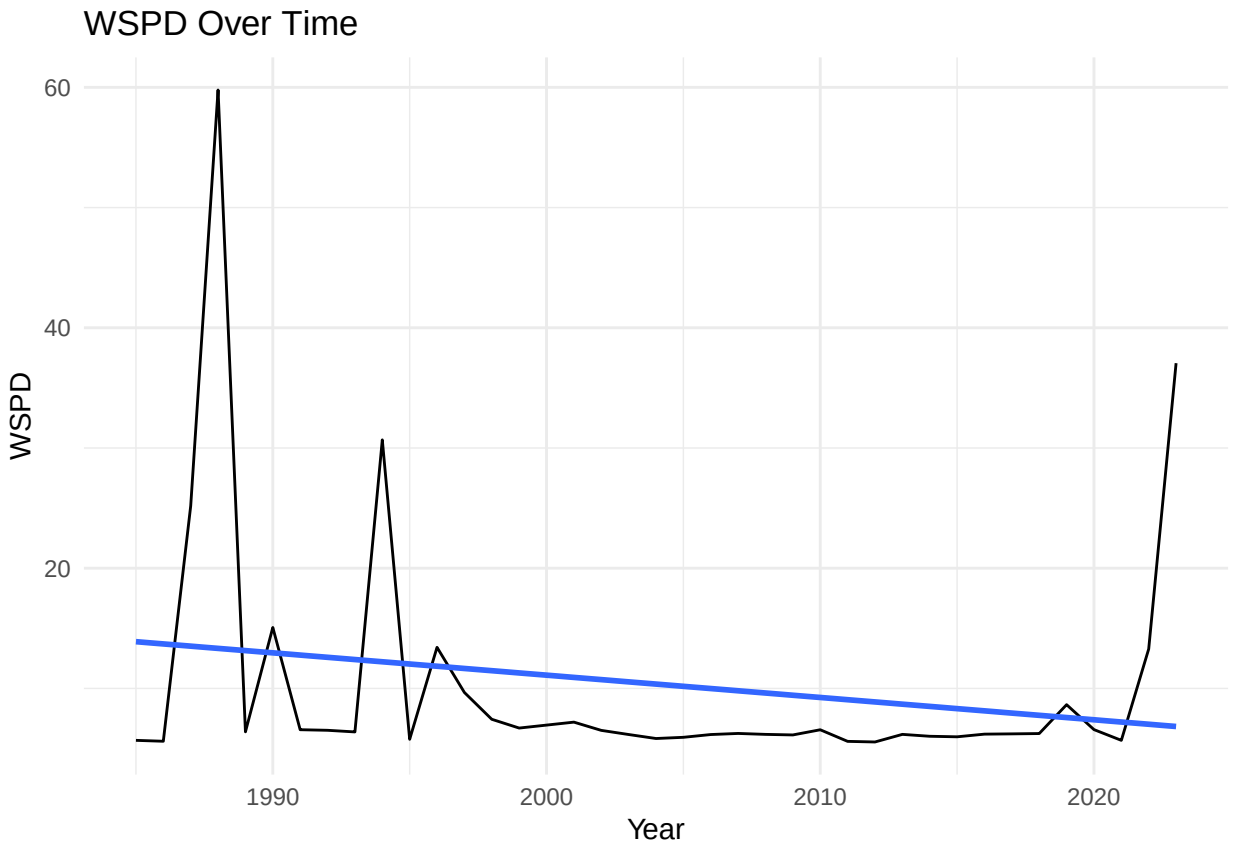
```
## Warning: Removed 21 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

```
## Warning: Removed 21 rows containing missing values or values outside the scale range
## (`geom_line()`).
```




```
ggplot(yearly_trends, aes(x = Year, y = avg_WSPD)) +
  geom_line() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "WSPD Over Time",
       x = "Year",
       y = "WSPD") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

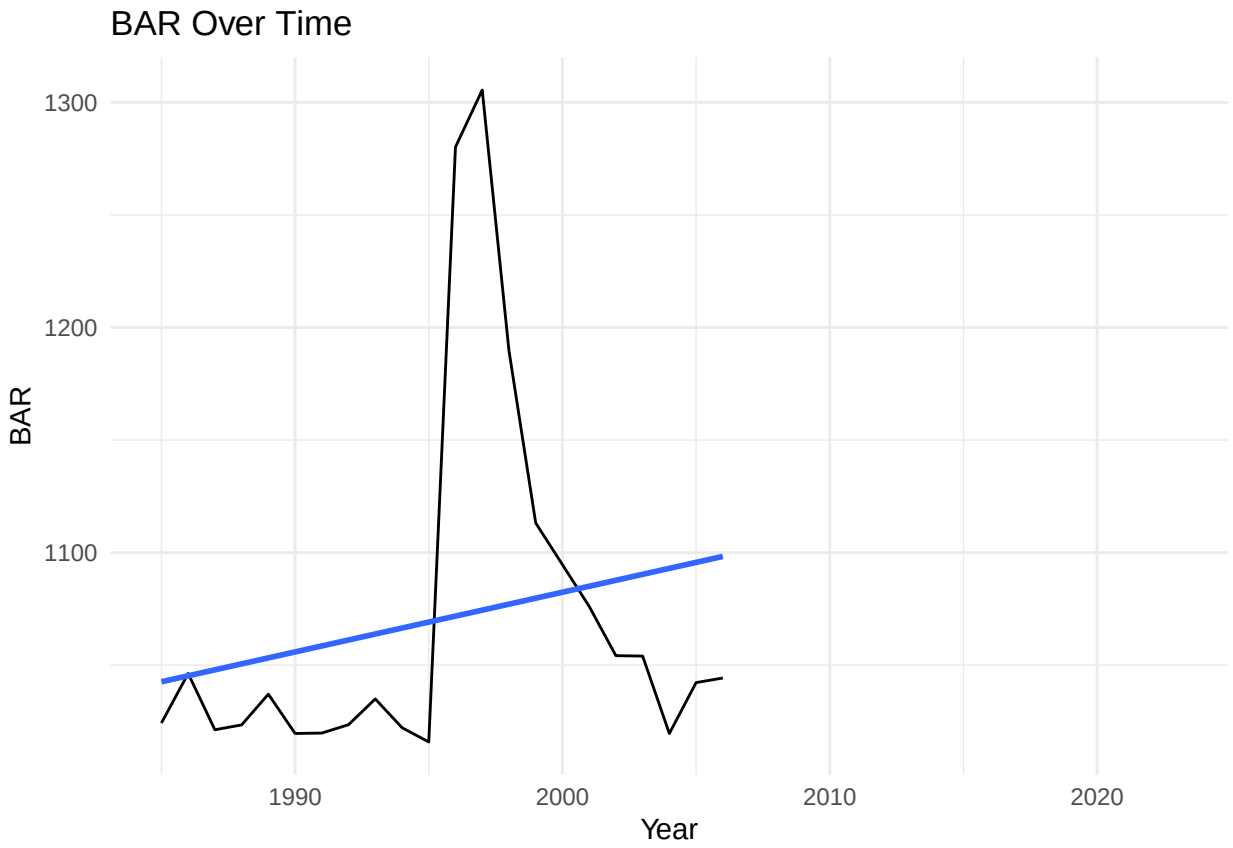



```
ggplot(yearly_trends, aes(x = Year, y = avg_BAR)) +
  geom_line() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "BAR Over Time",
       x = "Year",
       y = "BAR") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 17 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

```
## Warning: Removed 17 rows containing missing values or values outside the scale range
## (`geom_line()`).
```

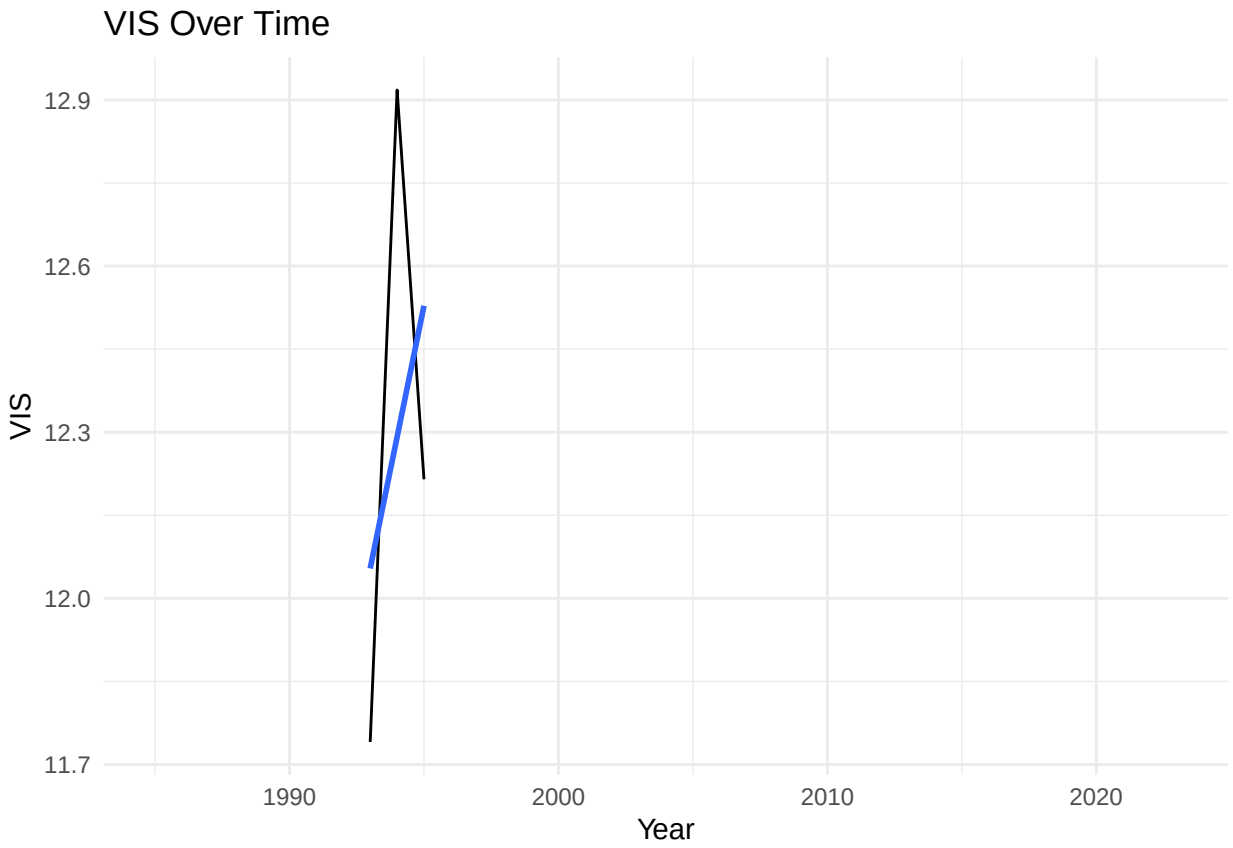



```
ggplot(yearly_trends, aes(x = Year, y = avg_VIS)) +
  geom_line() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "VIS Over Time",
       x = "Year",
       y = "VIS") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 35 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

```
## Warning: Removed 35 rows containing missing values or values outside the scale range
## (`geom_line()`).
```

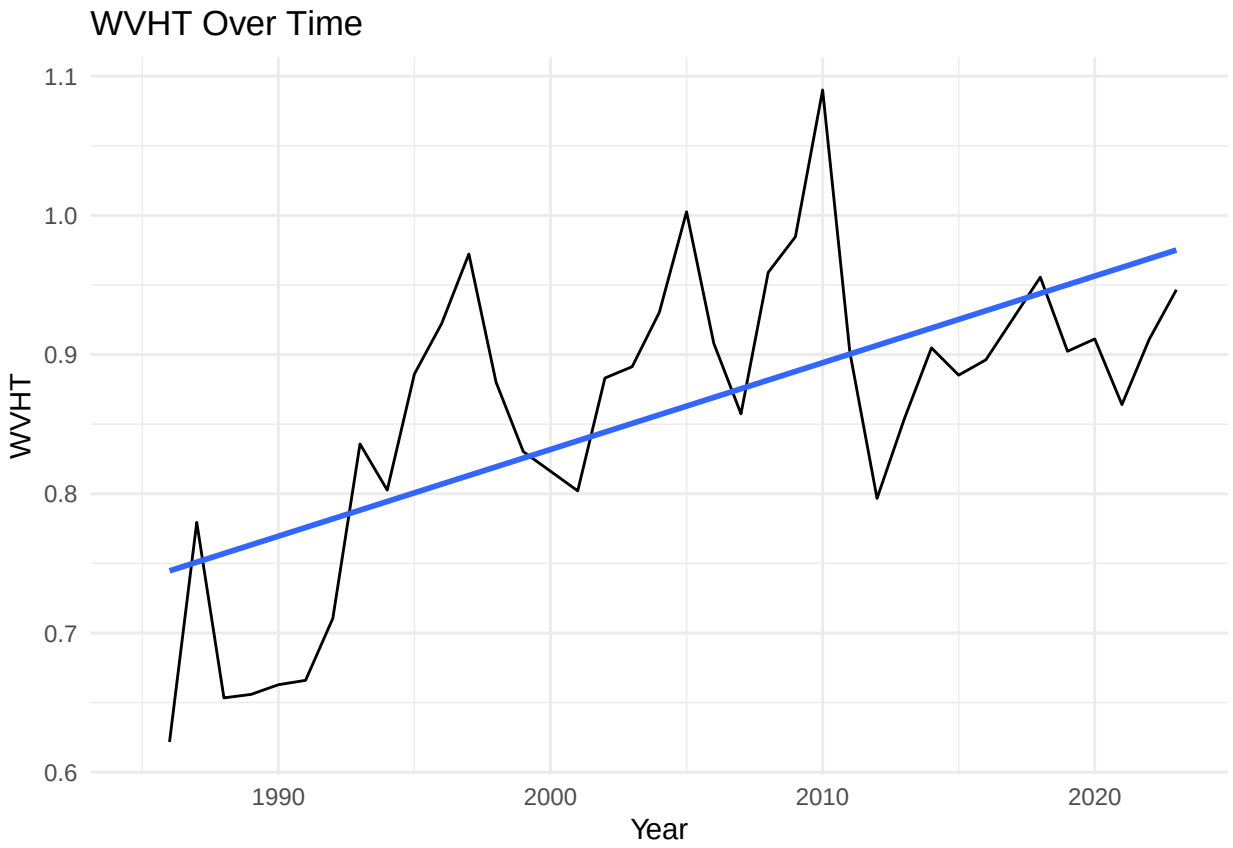



```
ggplot(yearly_trends, aes(x = Year, y = avg_WVHT)) +
  geom_line() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "WVHT Over Time",
       x = "Year",
       y = "WVHT") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_smooth()`).
```

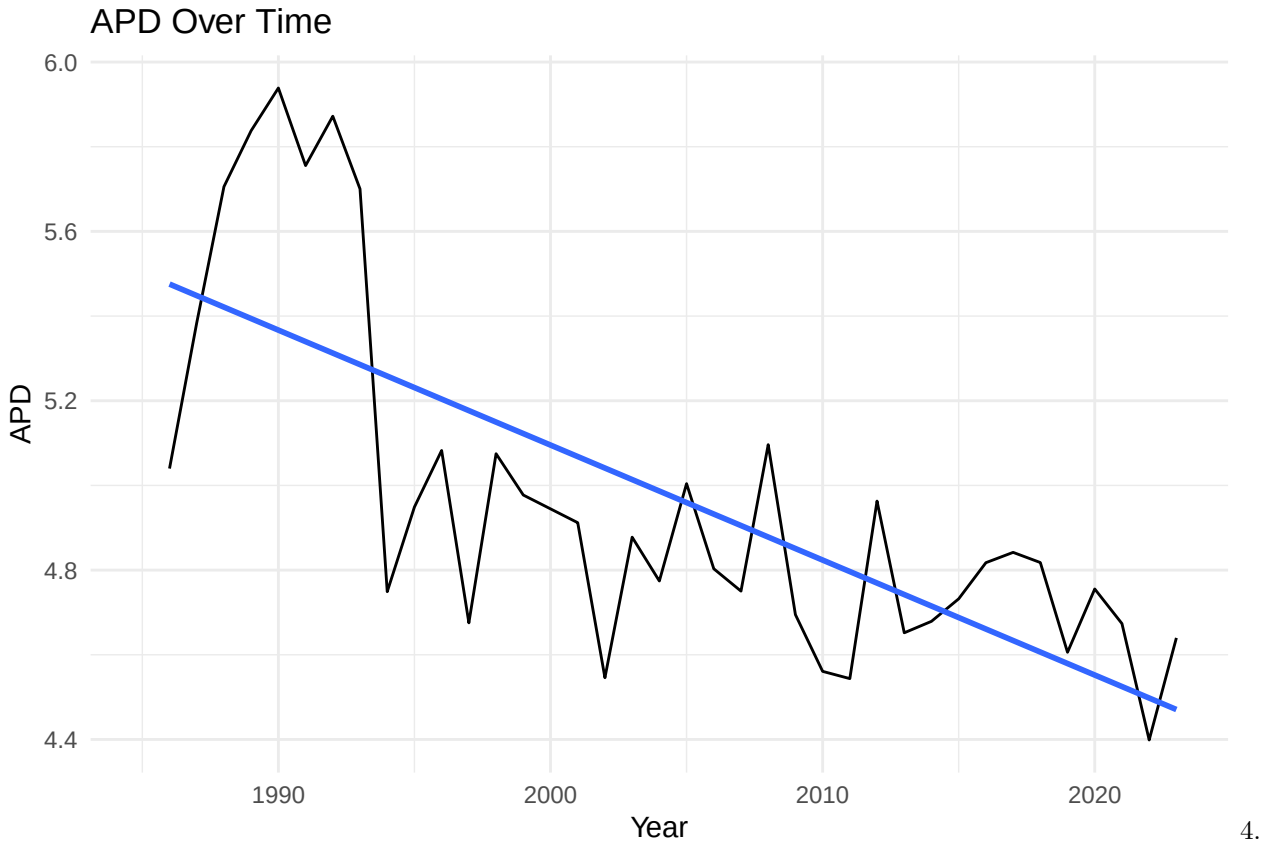
```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_line()`).
```




```
ggplot(yearly_trends, aes(x = Year, y = avg_APD)) +
  geom_line() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "APD Over Time",
       x = "Year",
       y = "APD") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 1 row containing non-finite outside the scale range (`stat_smooth()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_line()`).
```



4. The first bracket will answer two first two steps, summary the rainfall data and make a visualization of yealy data.

```
raindata <- read.csv("RainFall.csv")
head(raindata)
```

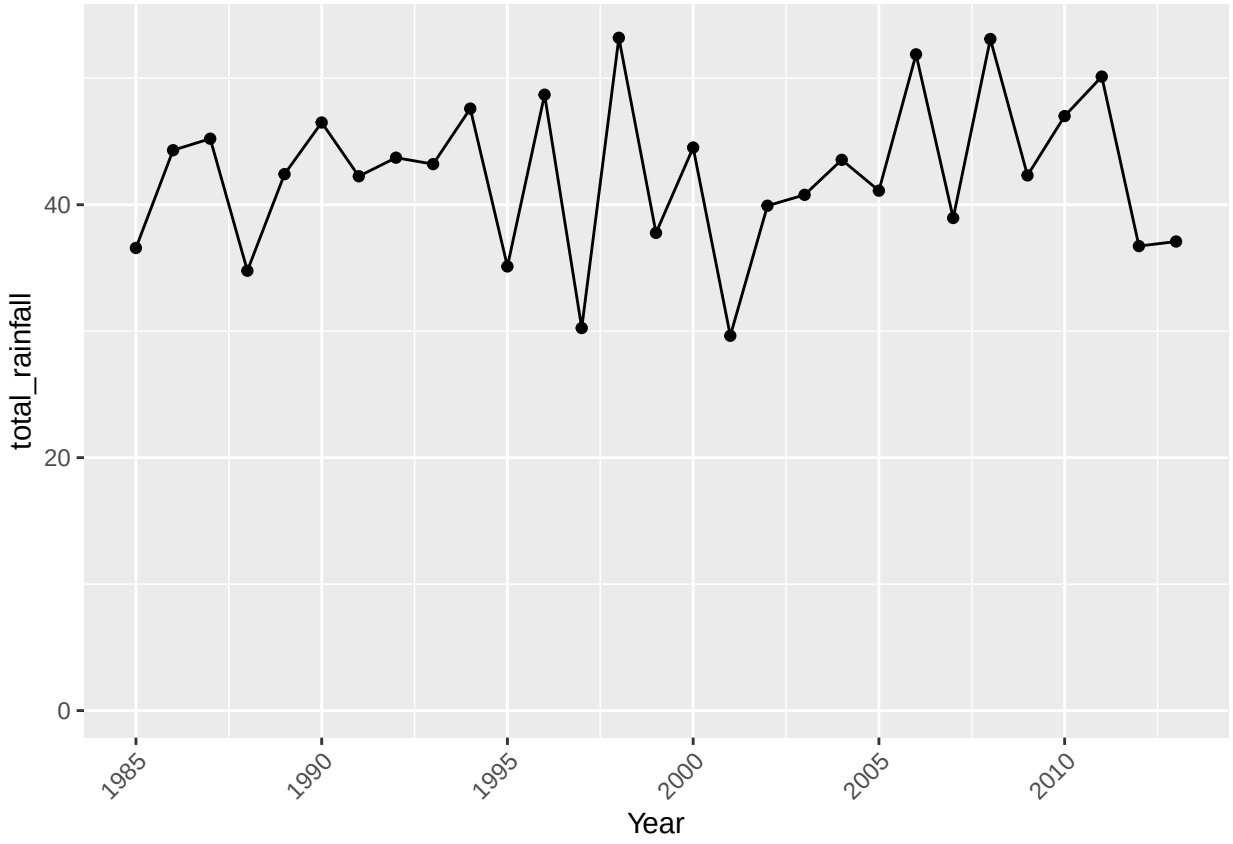
```
##       STATION                                STATION_NAME           DATE HPCP
## 1 COOP:190770 BOSTON LOGAN INTERNATIONAL AIRPORT MA US 19850101 01:00 0.00
## 2 COOP:190770 BOSTON LOGAN INTERNATIONAL AIRPORT MA US 19850101 09:00 0.01
## 3 COOP:190770 BOSTON LOGAN INTERNATIONAL AIRPORT MA US 19850101 10:00 0.01
## 4 COOP:190770 BOSTON LOGAN INTERNATIONAL AIRPORT MA US 19850101 11:00 0.01
## 5 COOP:190770 BOSTON LOGAN INTERNATIONAL AIRPORT MA US 19850101 12:00 0.01
## 6 COOP:190770 BOSTON LOGAN INTERNATIONAL AIRPORT MA US 19850101 13:00 0.01
##   Measurement.Flag Quality.Flag
## 1                g           NA
## 2                            NA
## 3                            NA
## 4                            NA
## 5                            NA
## 6                            NA
```

```
#Calculating the yearly data of rainfall and make a visualization.

raindata <- raindata %>%
  mutate(DATE = as.POSIXct(DATE, format = "%Y%m%d %H:%M"),
         Year = as.numeric(format(DATE, "%Y")))

yearly_rainfall <- raindata %>%
  group_by(Year) %>%
```

```
  summarise(total_rainfall = sum(HPCP, na.rm = TRUE)) %>%
  arrange(Year)

ggplot(yearly_rainfall, aes(x = Year, y = total_rainfall)) +
  geom_point()+
  geom_line()+
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1)
    )
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_line()`).
```




```
summary(yearly_rainfall)
```

```
##       Year      total_rainfall
##  Min.   :1985   Min.   : 0.50
##  1st Qu.:1992   1st Qu.:37.26
##  Median :1999   Median :42.37
##  Mean   :1999   Mean   :40.96
##  3rd Qu.:2006   3rd Qu.:46.18
##  Max.   :2013   Max.   :53.20
##  NA's   :1
```

The second bracket will create a simple model

```r
head(combined_data)
```

```
##       YY    MM    DD    hh    WD  WSPD   GST  WVHT   DPD   APD   MWD    BAR
##    <int> <int> <int> <int> <int> <num> <num> <num> <num> <num> <int>  <num>
## 1:    85     1     1     1    80     4     5    NA    NA    NA    NA 1030.0
## 2:    85     1     1     2   100     4     5    NA    NA    NA    NA 1030.1
## 3:    85     1     1     3   100     4     5    NA    NA    NA    NA 1029.4
## 4:    85     1     1     4   110     4     5    NA    NA    NA    NA 1028.6
## 5:    85     1     1     5    90     4     5    NA    NA    NA    NA 1027.8
## 6:    85     1     1     6    60     4     6    NA    NA    NA    NA 1027.7
##     ATMP  WTMP  DEWP   VIS  Year  YYYY  TIDE    mm   #YY  WDIR  PRES
##    <num> <num> <num> <num> <int> <int> <num> <int> <int> <int> <num>
## 1:   5.1   6.7    NA    NA  1985    NA    NA    NA    NA    NA    NA
## 2:   5.6   6.6    NA    NA  1985    NA    NA    NA    NA    NA    NA
## 3:   5.8   6.7    NA    NA  1985    NA    NA    NA    NA    NA    NA
## 4:   5.8   6.7    NA    NA  1985    NA    NA    NA    NA    NA    NA
## 5:   5.3   6.7    NA    NA  1985    NA    NA    NA    NA    NA    NA
## 6:   5.5   6.7    NA    NA  1985    NA    NA    NA    NA    NA    NA
```

```r
yearly_data <- merge(yearly_rainfall, yearly_trends, by = "Year")
simple_model <- lm(total_rainfall ~ avg_ATMP + avg_WSPD + avg_PRES + avg_WVHT+avg_APD, data = yearly_da
summary(simple_model)
```

```
##
## Call:
## lm(formula = total_rainfall ~ avg_ATMP + avg_WSPD + avg_PRES +
##     avg_WVHT + avg_APD, data = yearly_data)
##
## Residuals:
##      22      23      24      25      26      27      28
## -1.5089  0.7234 -2.1951  0.8707  0.3950  0.1399  1.5749
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -432.1658   166.7007  -2.592    0.234
## avg_ATMP      -0.4675     1.3403  -0.349    0.786
## avg_WSPD     -28.8579     8.2997  -3.477    0.178
## avg_PRES       0.6284     0.2240   2.805    0.218
## avg_WVHT     124.1227    26.9713   4.602    0.136
## avg_APD      -22.2693    14.7372  -1.511    0.372
##
## Residual standard error: 3.321 on 1 degrees of freedom
##   (21 observations deleted due to missingness)
## Multiple R-squared:  0.9571, Adjusted R-squared:  0.7428
## F-statistic: 4.465 on 5 and 1 DF,  p-value: 0.344
```