

# STAT 451 - Visualizing Data - Autumn 2025

Ariane Ducellier

10/07/2025

Load R packages

```
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    3.5.2      v tibble    3.3.0
## v lubridate  1.9.4      v tidyr     1.3.1
## v purrr      1.1.0
```

```
## -- 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
```

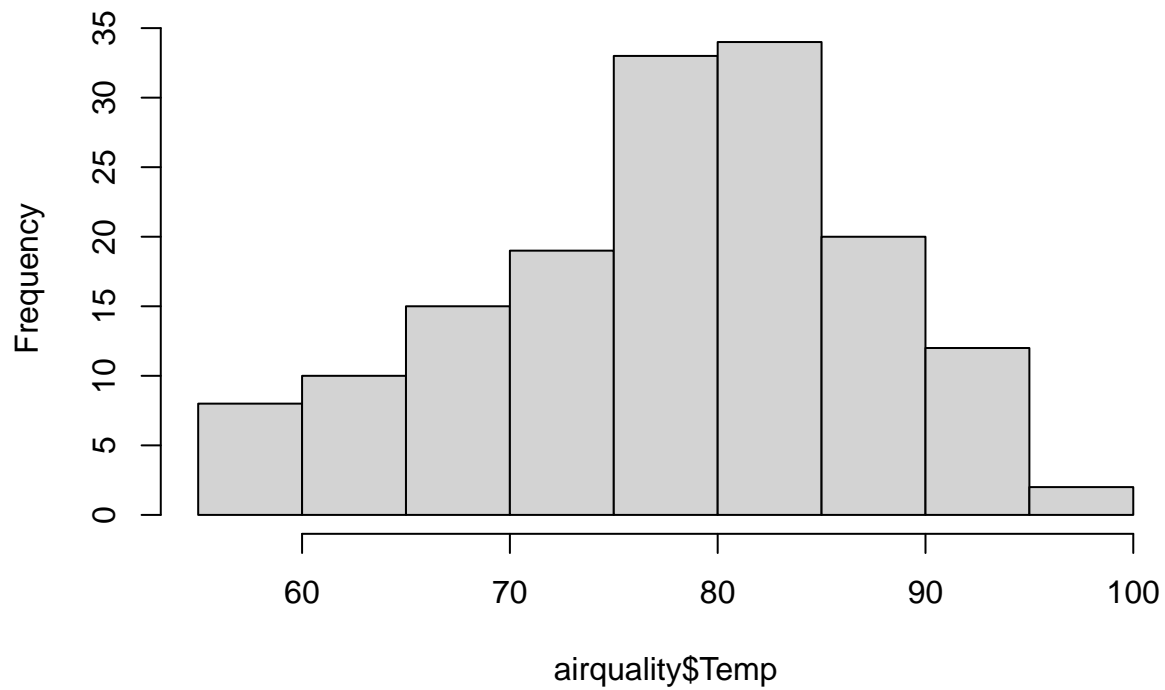
```
library(Lock5Data)
```

## Part 1 - Basic Plotting in ggplot2

### Histograms

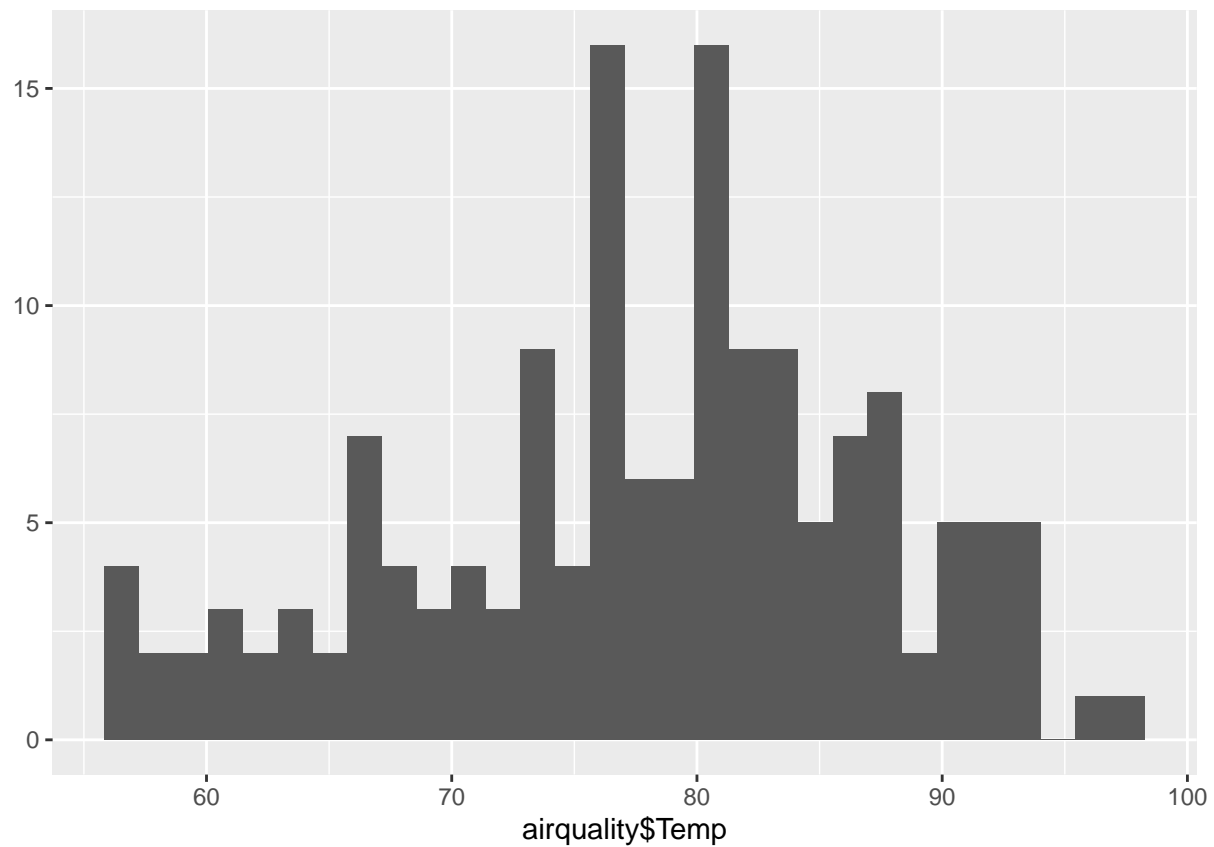
```
hist(airquality$Temp)
```

## Histogram of airquality\$Temp



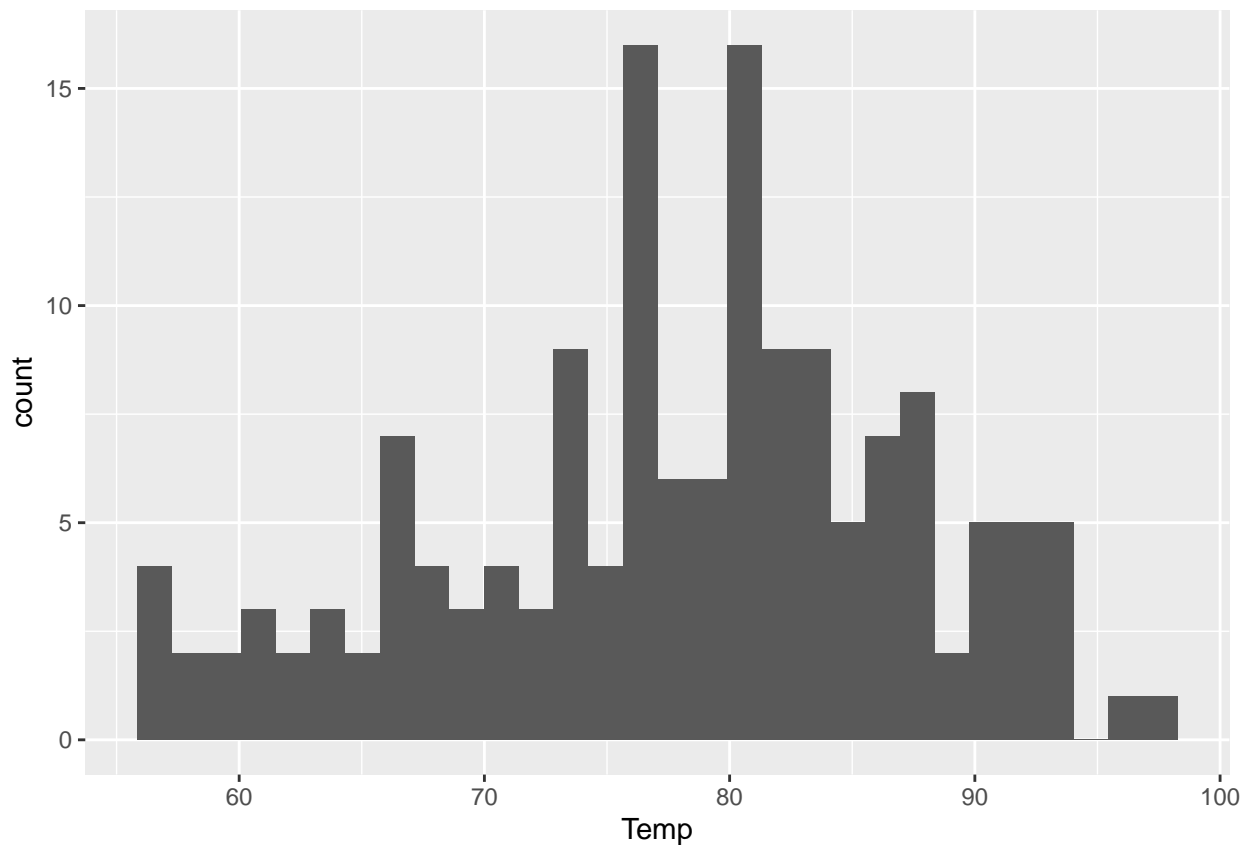
```
qplot(airquality$Temp)
```

```
## Warning: `qplot()` was deprecated in ggplot2 3.4.0.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was  
## generated.  
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
ggplot(airquality, aes(x=Temp)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

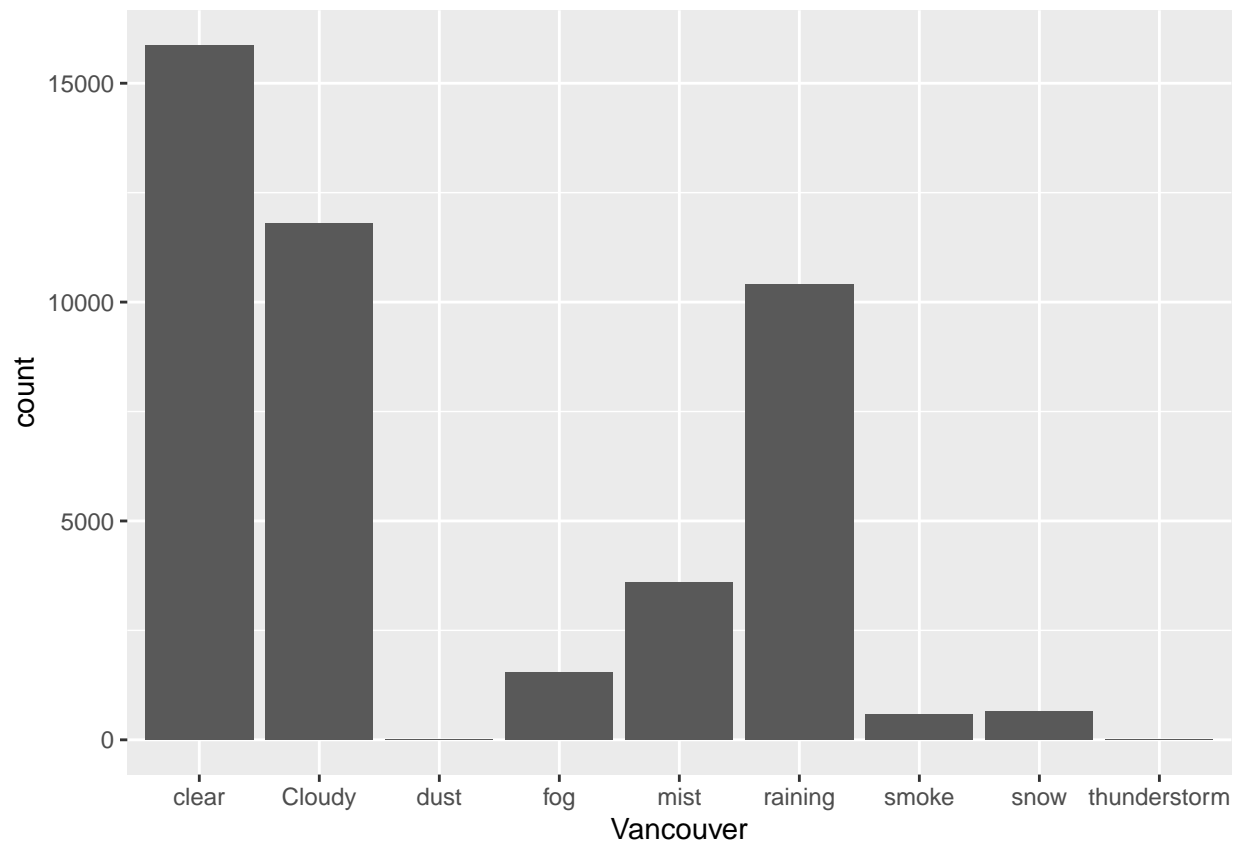


## Bar plots

```
df_desc <- read_csv("../data/historical-hourly-weather-data/weather_description.csv")
```

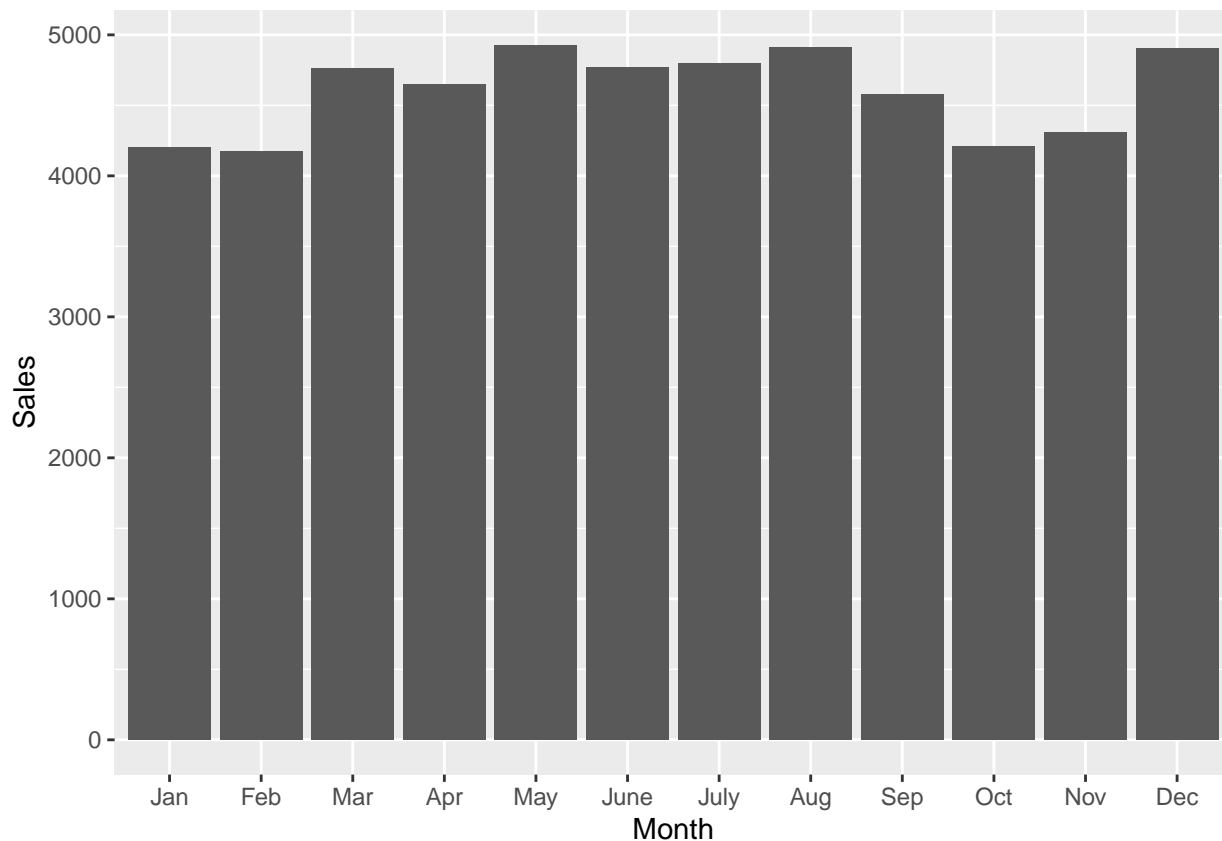
```
## New names:
## Rows: 44459 Columns: 4
## -- Column specification
## ----- Delimiter: "," chr
## (3): Vancouver, Seattle, San.Francisco dbl (1): ...1
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

```
ggplot(df_desc, aes(x=Vancouver)) + geom_bar()
```



```
df <- na.omit(RetailSales)
months_of_the_year <- c("Jan", "Feb", "Mar", "Apr", "May", "June",
                        "July", "Aug", "Sep", "Oct", "Nov", "Dec")

ggplot(df) +
  geom_bar(aes(x=factor(Month, months_of_the_year), y=Sales), stat="identity") +
  xlab("Month")
```

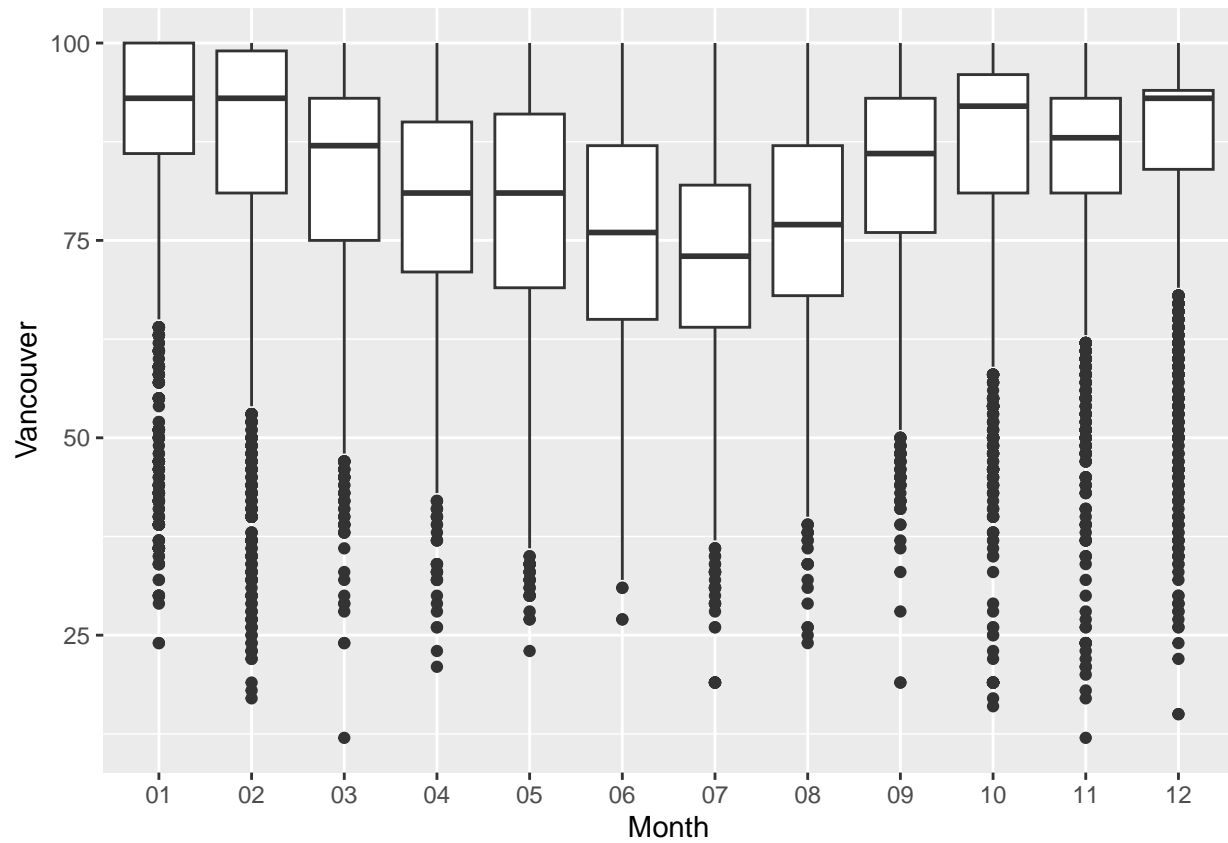


## Box plots

```
df_hum <- read_csv("../data/historical-hourly-weather-data/humidity.csv")

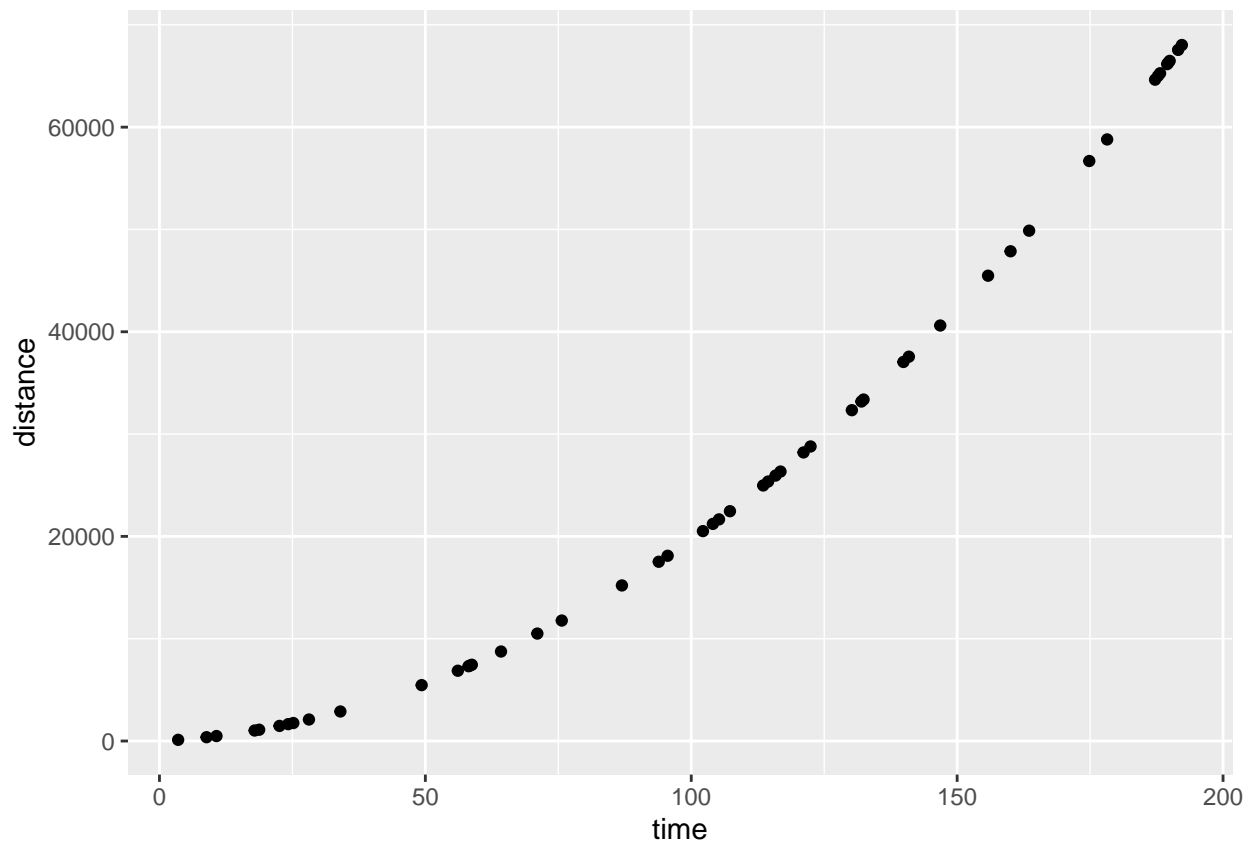
## Rows: 45253 Columns: 37
## -- Column specification -----
## Delimiter: ","
## dbl   (36): Vancouver, Portland, San Francisco, Seattle, Los Angeles, San Die...
## dtm   (1): datetime
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
df_hum$datetime <- as.character(df_hum$datetime)
df_hum$Month <- substr(df_hum$datetime, 6, 7)
ggplot(df_hum, aes(x=Month, y=Vancouver)) +
  geom_boxplot()

## Warning: Removed 1826 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



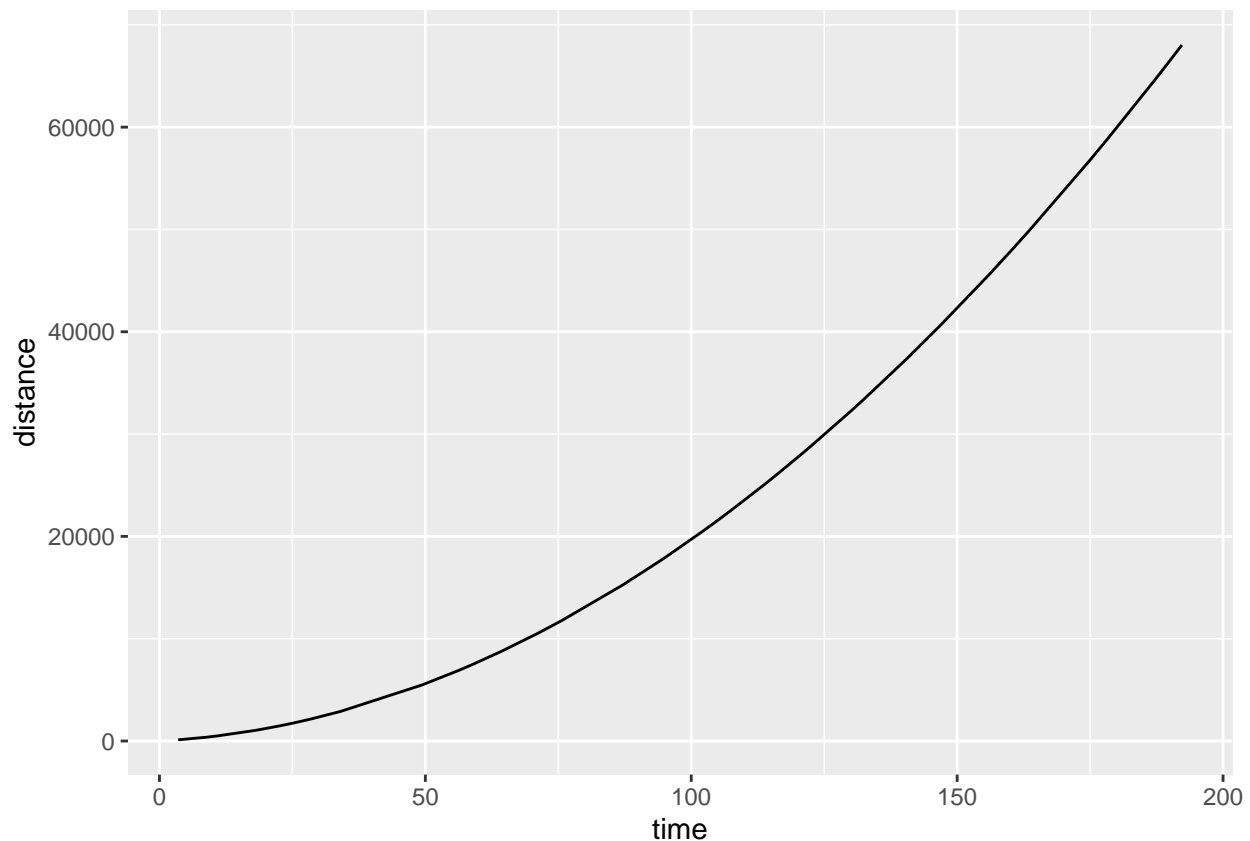
## Scatter plots and line plots

```
a = 3.4
v0 = 27
time <- runif(50, min=0, max=200)
distance <- sapply(time, function(x) v0 * x + 0.5 * a * x^2)
df <- data.frame(time,distance)
ggplot(df, aes(x=time, y=distance)) + geom_point()
```



```
ggplot(df, aes(x=time, y=distance)) + geom_line()
```



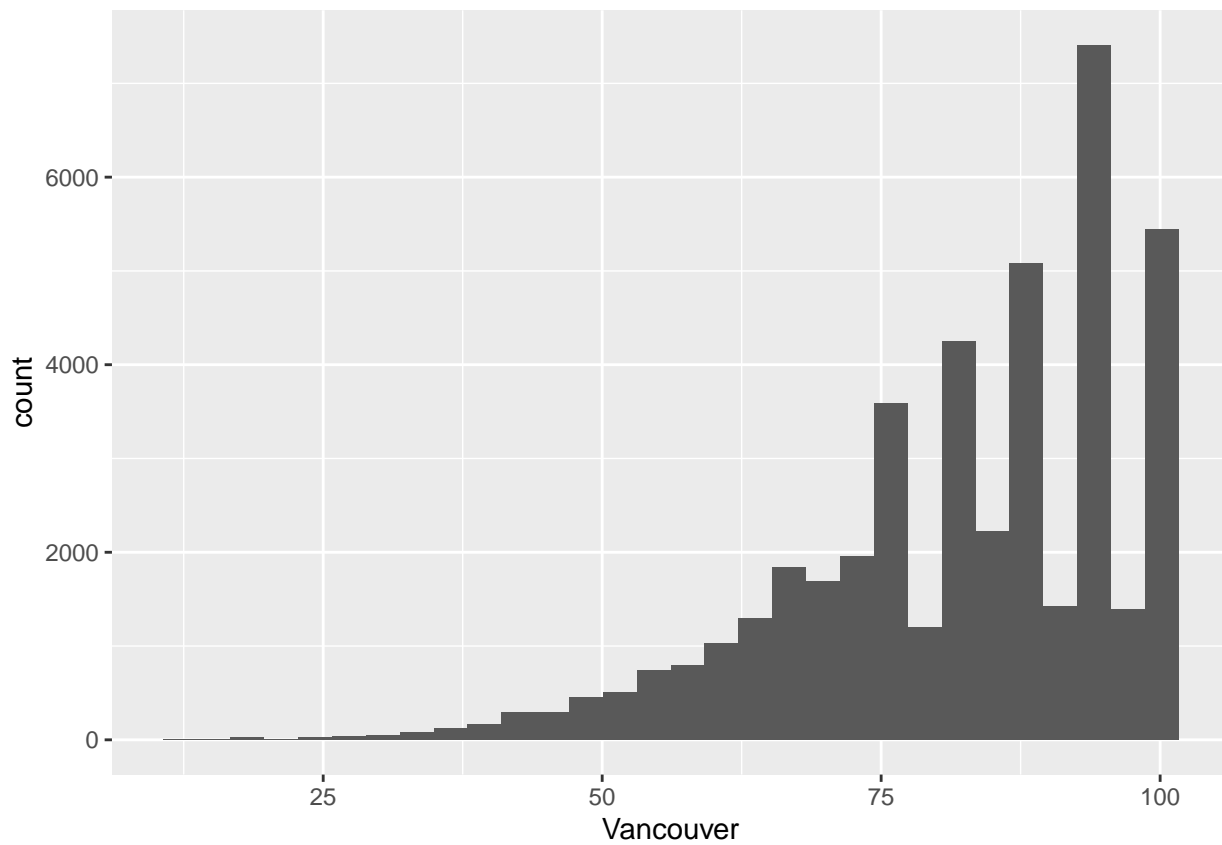


## Changing histogram defaults and adding aesthetics

```
df_hum <- read_csv("../data/historical-hourly-weather-data/humidity.csv")

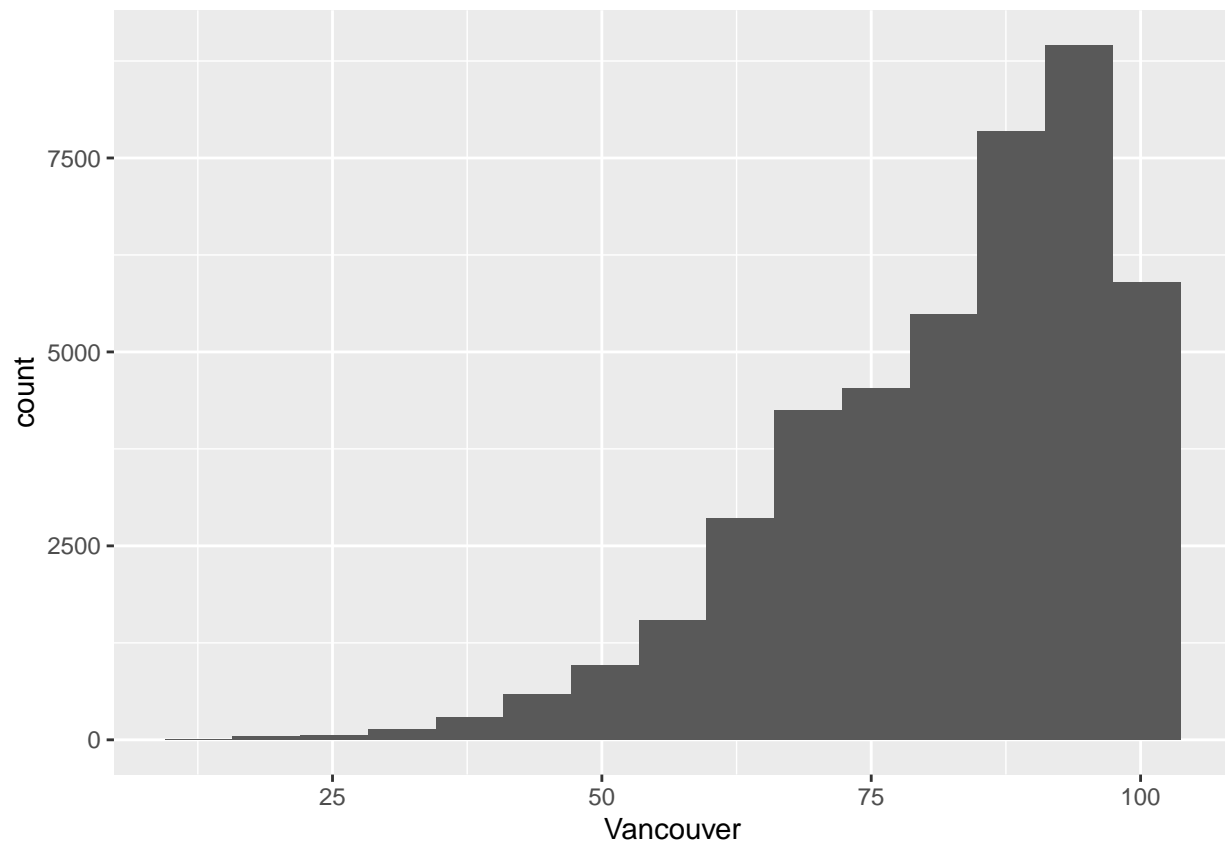
## Rows: 45253 Columns: 37
## -- Column specification -----
## Delimiter: ","
## dbl   (36): Vancouver, Portland, San Francisco, Seattle, Los Angeles, San Die...
## dtm   (1): datetime
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
ggplot(df_hum, aes(x=Vancouver)) + geom_histogram()

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 1826 rows containing non-finite outside the scale range
## (`stat_bin()`).
```



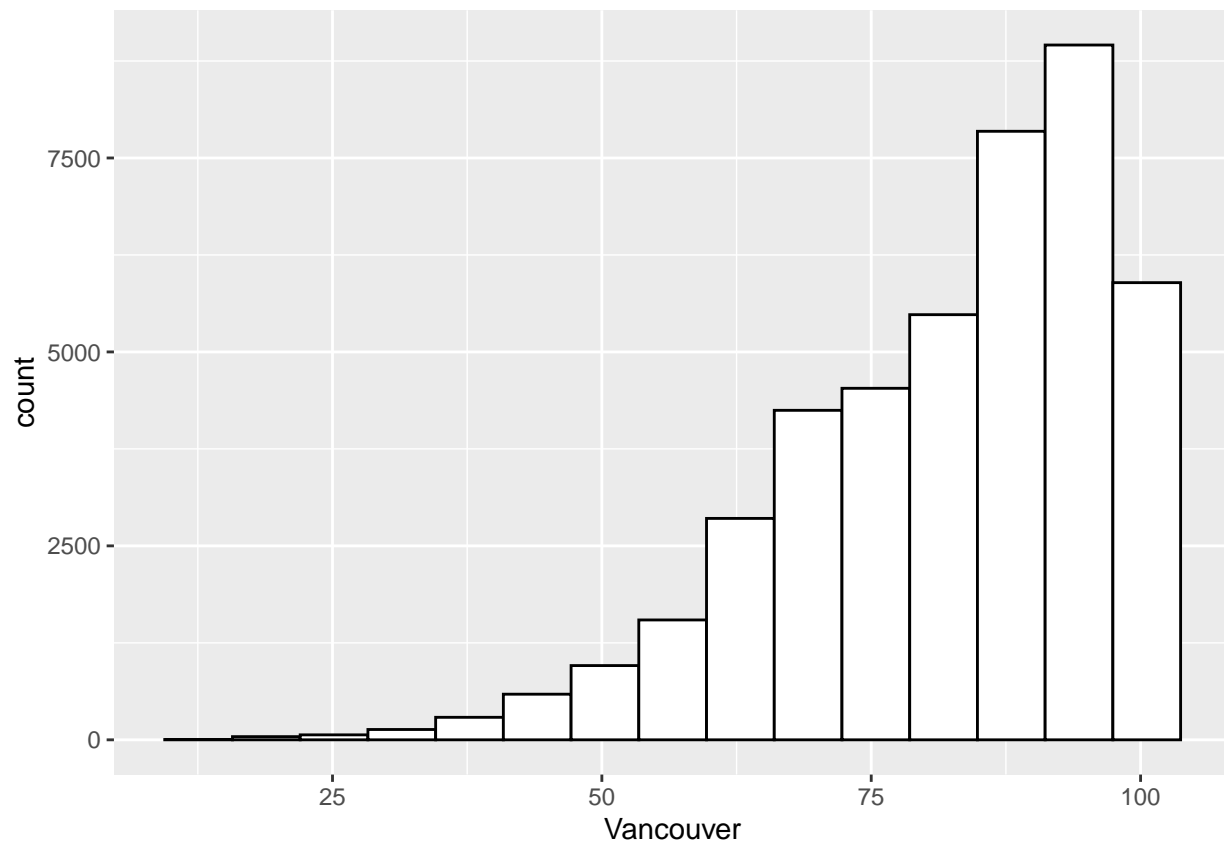
```
ggplot(df_hum, aes(x=Vancouver)) + geom_histogram(bins=15)
```

```
## Warning: Removed 1826 rows containing non-finite outside the scale range  
## (`stat_bin()`).
```



```
ggplot(df_hum, aes(x=Vancouver)) + geom_histogram(bins=15, fill="white", color=1)
```

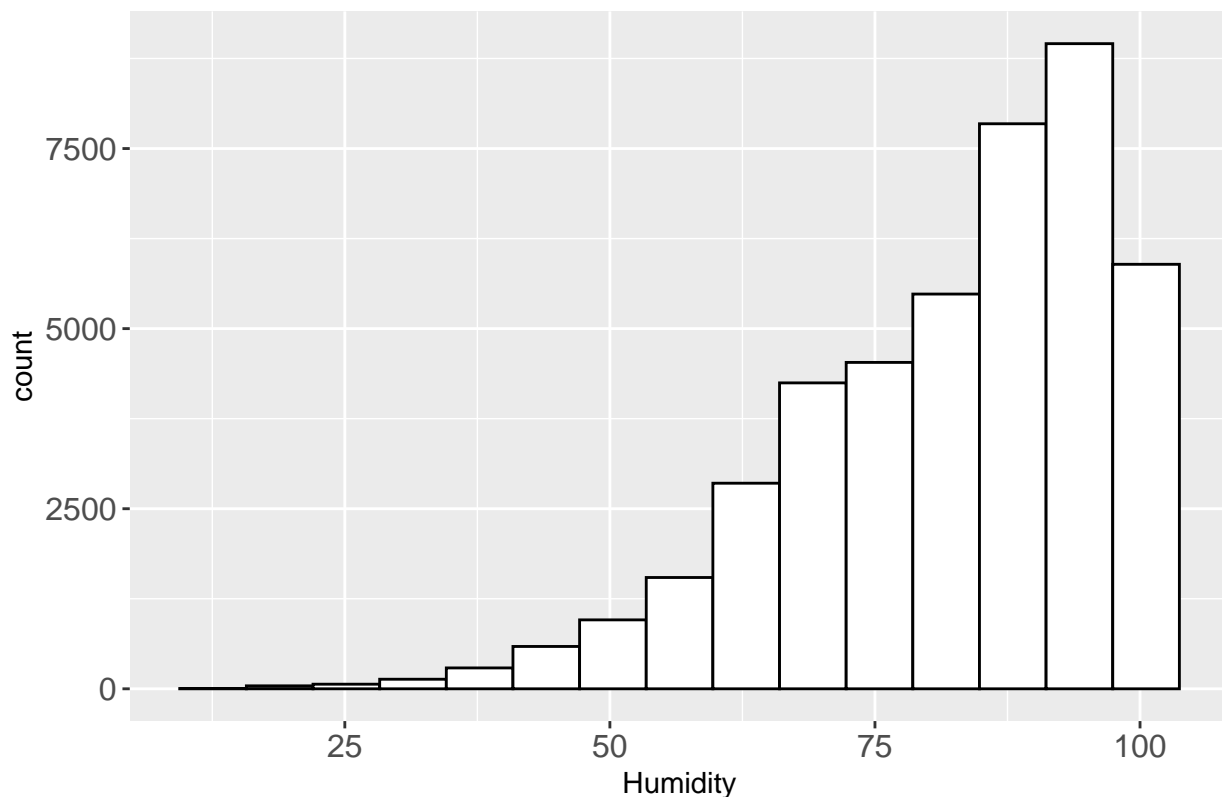
```
## Warning: Removed 1826 rows containing non-finite outside the scale range  
## (`stat_bin()`).
```



```
ggplot(df_hum, aes(x=Vancouver)) +  
  geom_histogram(bins=15, fill="white", color=1) +  
  ggtitle("Humidity for Vancouver city") +  
  xlab("Humidity") +  
  theme(axis.text.x=element_text(size=12), axis.text.y=element_text(size=12))
```

```
## Warning: Removed 1826 rows containing non-finite outside the scale range  
## (`stat_bin()`).
```

Humidity for Vancouver city



## Changing boxplot defaults and adding aesthetics

```
df_hum <- read_csv("../data/historical-hourly-weather-data/humidity.csv")

## Rows: 45253 Columns: 37
## -- Column specification -----
## Delimiter: ","
## dbl   (36): Vancouver, Portland, San Francisco, Seattle, Los Angeles, San Die...
## dtm   (1): datetime
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

df_hum$datetime <- as.character(df_hum$datetime)
df_hum$Month <- substr(df_hum$datetime, 6, 7)
ggplot(df_hum, aes(x=Month, y=Vancouver)) +
  geom_boxplot(color="gray20", fill="cadetblue2") +
  ylab("Humidity") +
  theme(axis.text.x=element_text(size=15),
        axis.text.y=element_text(size=15),
        axis.title.x=element_text(size=15, color="gray20"),
        axis.title.y=element_text(size=15, color="gray20"))

## Warning: Removed 1826 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```

