VIDO BIOINFO Bioinformatics and Systems Biology Lab

Canadian Bioinformatics Hub

Canadian Bioinformatics Workshops

Introduction to R Programming for Bioinformatics

Day 1- Module 2B: R Basic Plotting and Introduction to ggplot2

Mohamed Helmy, PhD
Principal Scientist and Adjunct Professor
Bioinformatics and Systems Biology Lab
VIDO, University of Saskatchewan

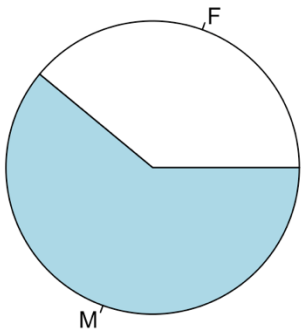
6-7 October 2025, VIDO, Saskatoon

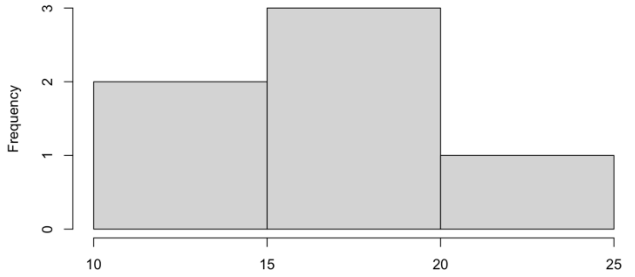
Why Data Import Matters?

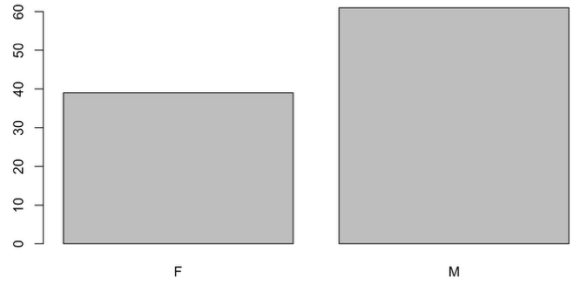
- Visualization
  - Essential for exploring and communicating data
  - Helps detect patterns, trends, and outliers
  - Translates raw numbers into insightful stories
  - Example in health data:
    - Age distribution of patients
    - Blood Pressure trends across groups

Base R Plotting Functions

- Built-in plotting system in R
- Common functions:
  - plot() → scatterplots, line plots
  - hist() → histograms
  - boxplot() → compare groups
  - barplot() → bar charts
  - pie() → pie charts
- Quick and simple for small projects







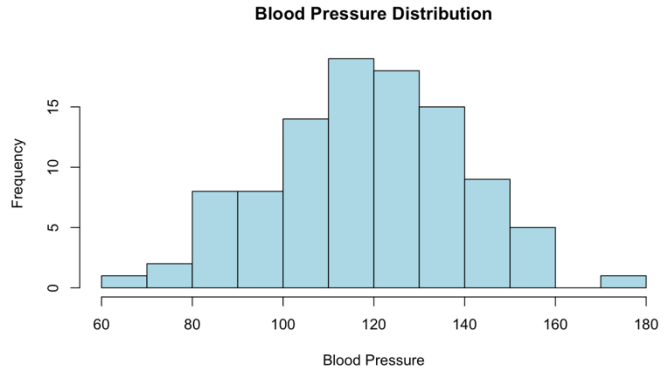
Hands-on: Base R Plots

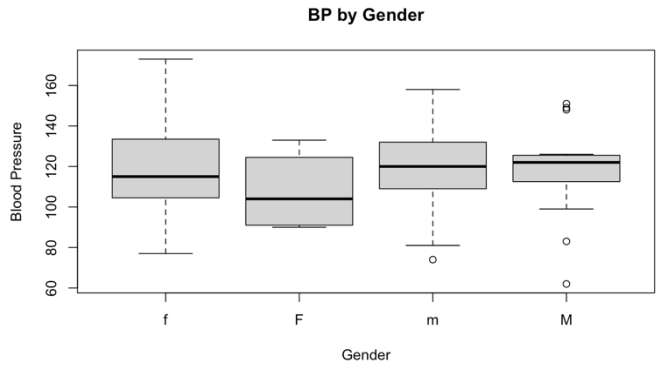
- **Histogram of Blood Pressure**
  - hist()
  - Tip: use *col="lightblue"* to change the color

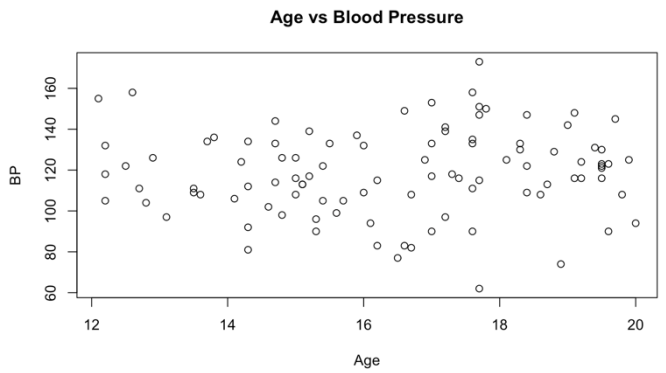
- **Boxplot of BP by Gender**
  - boxplot()
  - Tip: use *BloodPressure ~ Gender* to indicate the variables that you want to compare

- **Scatterplot Age vs BP**
  - plot()
  - Tip: use bp_dataAge, bp_dataBloodPressure to indicate the variables that you want to compare

General tips: *main*, *ylab*, *xlab* are for the plot title, y and x axis labels, respectively.







Hands-on: Base R Plots

- **Histogram of Blood Pressure**
  - hist()
  - Tip: use *col="lightblue"* to change the color

- **Boxplot of BP by Gender**
  - boxplot()
  - Tip: use *BloodPressure ~ Gender* to indicate the variables that you want to compare

- **Scatterplot Age vs BP**
  - plot()
  - Tip: use bp_dataAge, bp_dataBloodPressure to indicate the variables that you want to compare

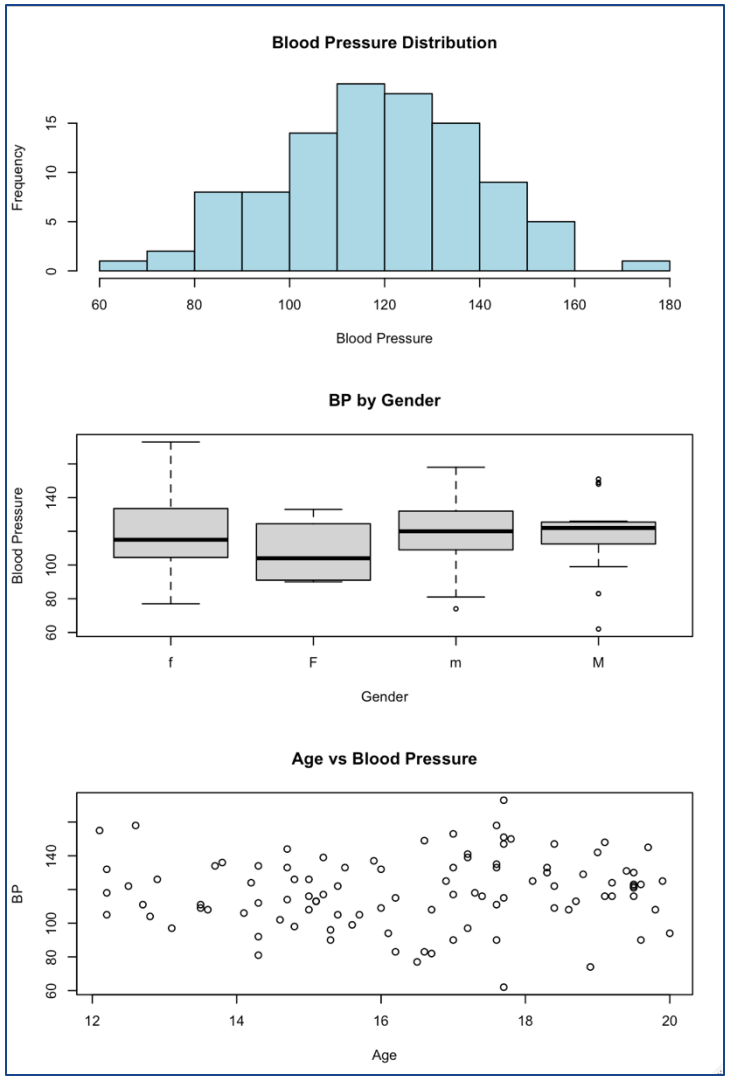
General tips: main, ylab, xlab are for the plot title, y and x axis labels, respectively.

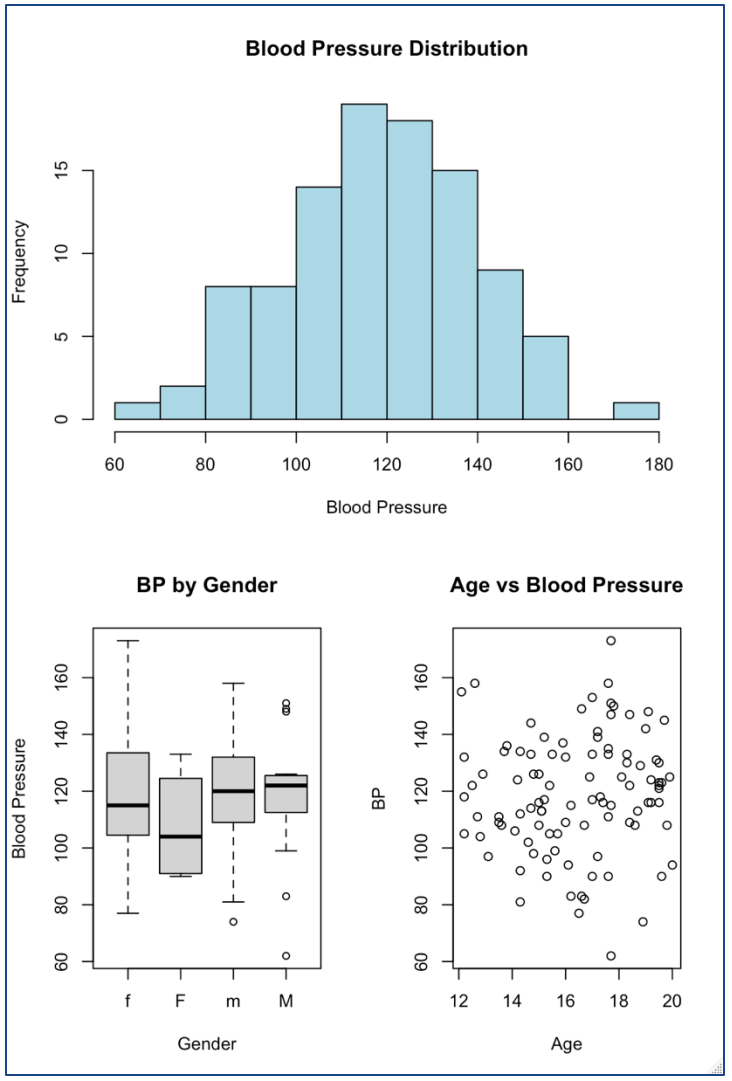
```r
# Basic plotting in R

# Histogram of Blood Pressure
hist(bp_data$BloodPressure, main="Blood Pressure Distribution",
     xlab="Blood Pressure", col="lightblue")

# Boxplot of BP by Gender
boxplot(BloodPressure ~ Gender, data=bp_data,
        main="BP by Gender", xlab="Gender", ylab="Blood Pressure")

# Scatterplot Age vs BP
plot(bp_data$Age, bp_data$BloodPressure,
     main="Age vs Blood Pressure", xlab="Age", ylab="BP")
```

Multi-panel plots in base R

- **The par() function**
  - par(mfrow = c(rows, columns))
  - par(mfrow = c(3, 1))

- **The layout() function**
  - layout(matrix(c(1, 1, 2, 3), nrow = 2, byrow = T))



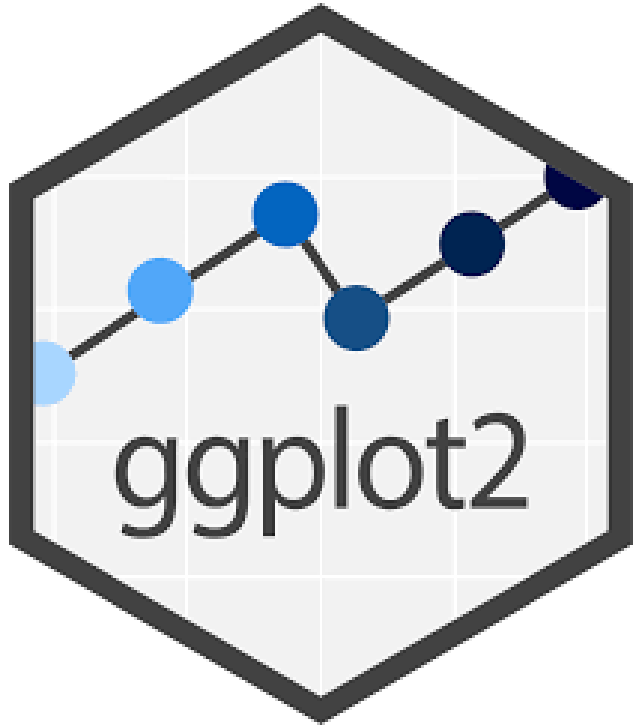


Limitations of Base R Graphics

- **Customization is limited** (fonts, colors, themes)
- **Complex plots = messy code**
- **Not consistent with modern data pipelines**
- **Useful for quick checks, but less for polished reports**

Introduction to ggplot2

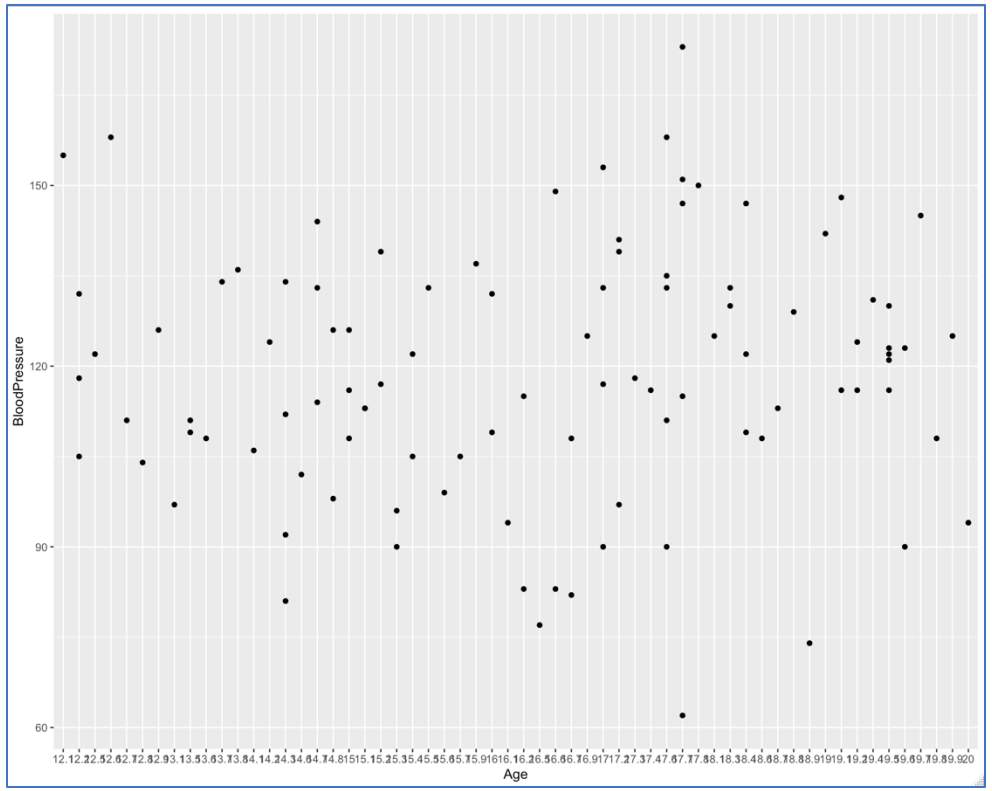
- Part of the *tidyverse* package collection
- Implements the *Grammar of Graphics*
- Flexible
- Powerful
- Widely used in bioinformatics
- Recommended resource: https://ggplot2.tidyverse.org/articles/ggplot2.html




```
# example syntax
ggplot(data, aes(x, y)) + geom_*()
```

ggplot2 Basics (Syntax)

```
library(ggplot2)
# Scatterplot: Age vs Blood Pressure
ggplot(bp_data, aes(x=Age, y=BloodPressure)) +
  geom_point()
```



- **ggplot() → declare data & aesthetics**
- **aes() → map variables (x, y, color, size)**
- **geom_point() → adds scatterplot layer**

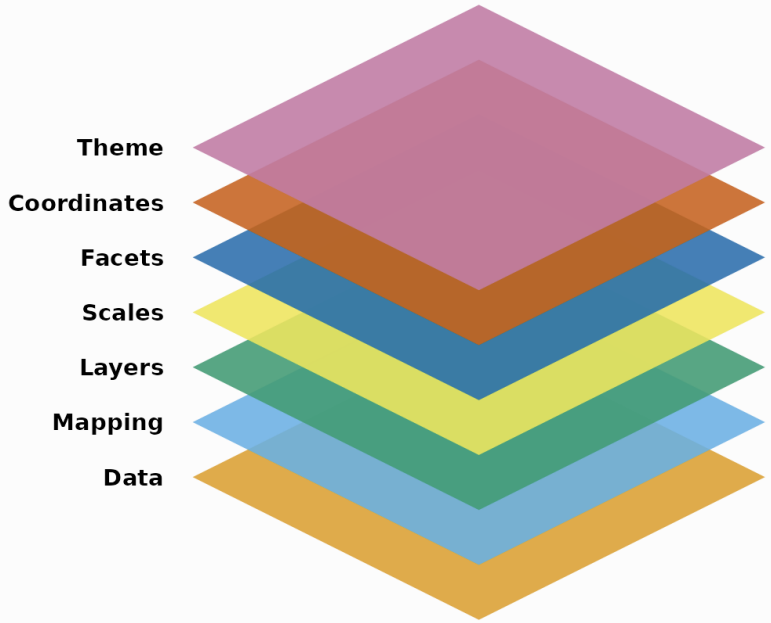
ggplot2 Basics (layers)

- **Data:**
  - The foundation of every graphic
  - ggplot stores the data to be used later by other parts of the plotting system.
  - ggplot(data = mpg)

- **Mapping:**
  - A set of instructions on how parts of the data are mapped onto aesthetic attributes of geometric objects.
  - A mapping can be made by using the aes() function
  - ggplot(bp_data, aes(x=Age, y=BloodPressure))

- **Layers:**
  - They take the mapped data and display it in something humans can understand
  - Every layer consists of three important parts:
    - The geometry: how data is displayed
    - The statistical transformation: what of the data is displayed.
    - The position adjustment: where a piece of data is displayed.

For structure, we go over the 7 composable parts that come together as a set of instructions on how to draw a chart.

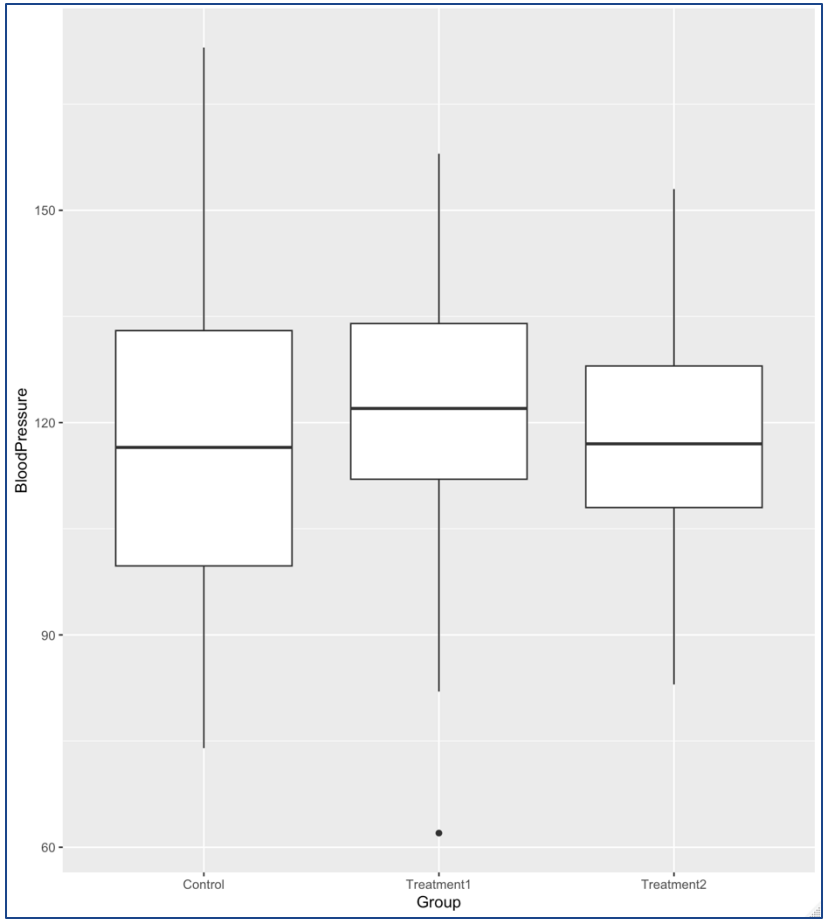


Source: https://ggplot2.tidyverse.org/articles/ggplot2.html

Common Geoms in ggplot2

- **geom_point() → scatterplots**
- **geom_histogram() → histograms**
- **geom_boxplot() → boxplots**
- **geom_bar() → bar charts**

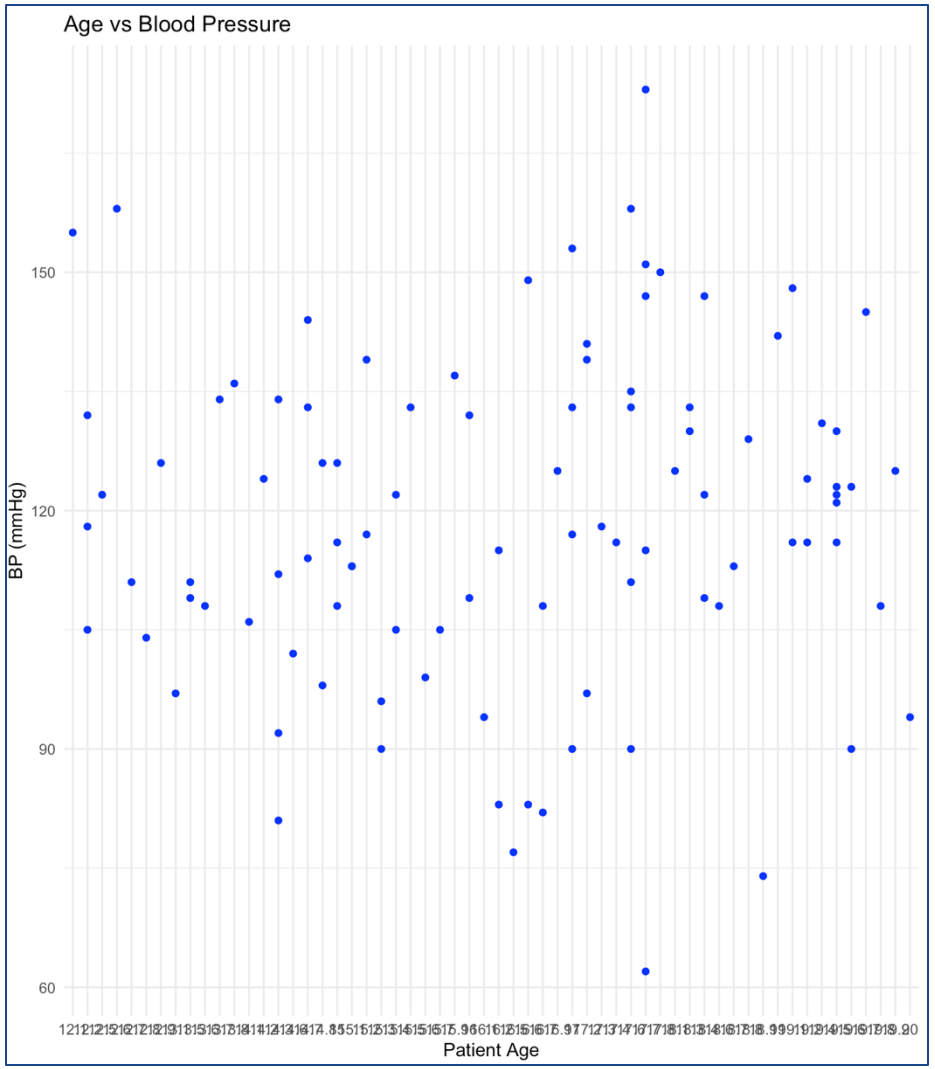
```
library(ggplot2)
# Scatterplot: Age vs Blood Pressure
ggplot(bp_data, aes(x=Age, y=BloodPressure)) +
  geom_point()

# boxplot
ggplot(bp_data, aes(x=Group, y=BloodPressure)) +
  geom_boxplot()
```



Customizing ggplot2 plots

- **Add titles, axis labels, themes**

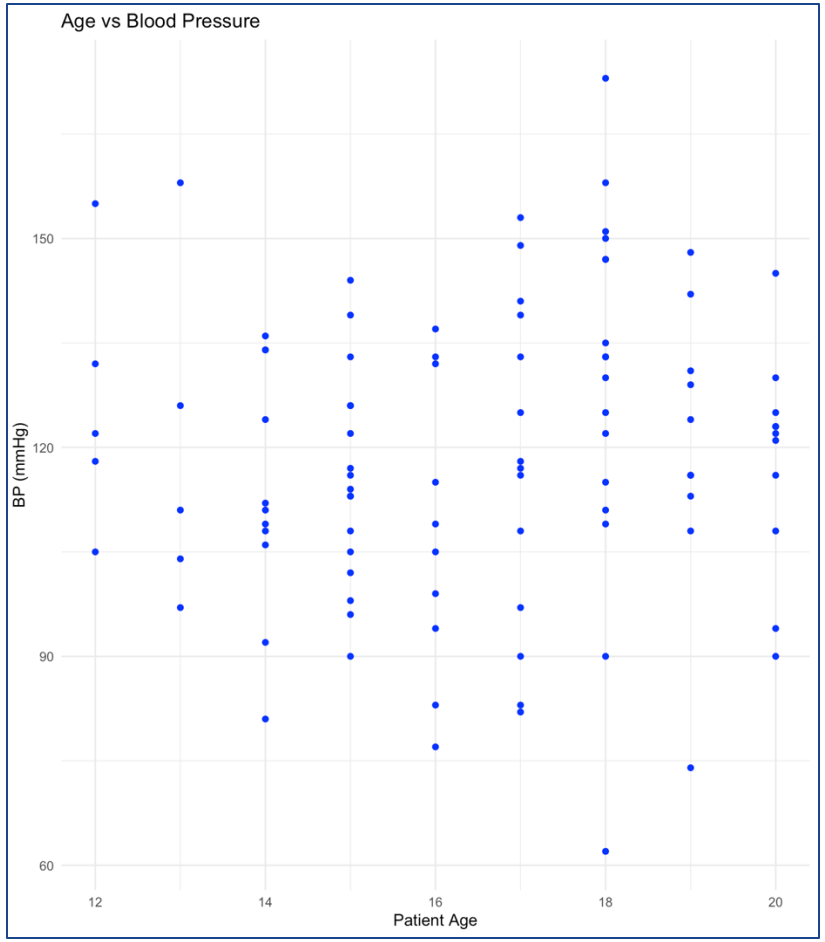
```
# Customizing ggplot2
ggplot(bp_data, aes(x = Age, y = BloodPressure)) +
  geom_point(color = "blue") +
  labs(
    title = "Age vs Blood Pressure",
    x = "Patient Age",
    y = "BP (mmHg)"
  ) +
  theme_minimal()
```



Customizing ggplot2 plots

- **Consistent, publication-quality graphics**

```r
# Customizing ggplot2
ggplot(bp_data, aes(x = round(as.numeric(Age), 0), y = BloodPressure)) +
  geom_point(color = "blue") +
  labs(
    title = "Age vs Blood Pressure",
    x = "Patient Age",
    y = "BP (mmHg)"
  ) +
  theme_minimal()
```



Saving Plots

- **Supports PNG, PDF, TIFF, JPEG**

- **Control width, height, resolution**

```r
# Save last plot as PNG
ggsave("Age_BP_Scatter.png", width=6, height=4)

# Save specific plot object
p <- ggplot(bp_data, aes(x=Age, y=BloodPressure)) +
  geom_point()
ggsave("data/scatter_plot.png", plot=p)
```

Hands-on: ggplot Examples

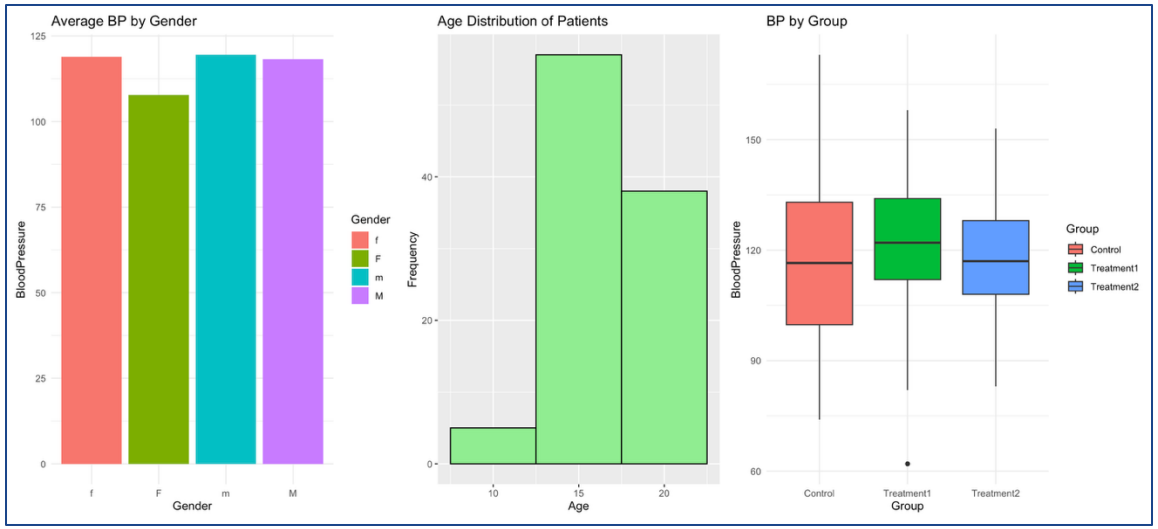
- **Create a bar plot of the number of patients in each Group (A, B, C).**
  - geom_bar()
  - Tip: use the *fill*="" and color="" parameters

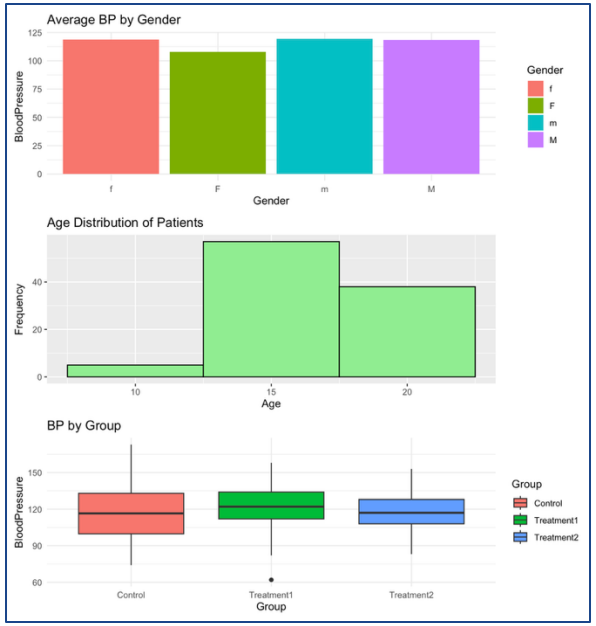
- **Create a histogram of Age and choose a reasonable binwidth.**
  - geom_histogram()
  - Tip: make sure the Age column is numeric not character

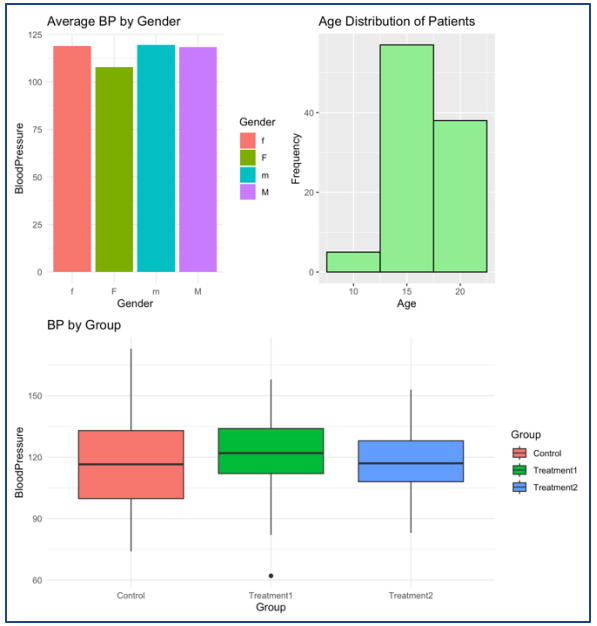
- **Make a scatterplot of Age vs BloodPressure, color points by Group, and add a title.**
  - geom_point()
  - Tip: use *labs (title="", x="", y="")* parameters

- **Create a single figure that shows the following three plots side by side (patchwork pakage):**
  - Bar plot: Average Blood Pressure by Gender
  - Histogram: Age Distribution of Patients
  - Boxplot: Blood Pressure by Group

Hands-on: ggplot Examples







Hands-on: ggplot Examples

- **Use the blood pressure dataset**
  - Read the file into R
  - Make sure all the entries of the “Date” column are in the YMD format
  - Create a new column and store the year in this column
  - Filter the patients based on the year and sex

```
# Work with date
library(readr)
library("lubridate")

# read ALL data
bp <-  read.csv2("Desktop/R/data/BloodPressure_wDates.csv", sep = ",")

# Convert date column and extract year
bp$Date <- ymd(bp$Date)
bp$Year <- year(bp$Date)

# Filtering blood pressure patients by year and gender
subset(bp, Year == 2003 & Gender == "f")
```

THANK YOU

VIDO

VACCINE AND INFECTIOUS DISEASE ORGANIZATION

VIDO.ORG

UNIVERSITY OF SASKATCHEWAN