Exploratory Data Analysis

PRACTICAL:03

>library(tidyverse)

* **Visualizing Distribution**

>ggplot(data = diabetes) + geom\_bar(mapping = aes(x = BMI))

> diabetes %>%+ count(BMI)

>ggplot(data = diabetes) + geom\_histogram(mapping = aes(x = Age), binwidth = 0.5)

>ggplot(data = diabetes, mapping = aes(x = BMI, color = Age)) + geom\_freqpoly(binwidth = 0.1)

* **Typical Values**

>ggplot(data = diabetes, mapping = aes(x = BMI)) + geom\_histogram(binwidth = 0.01)

>ggplot(data = faithful, mapping = aes(x = eruptions)) + geom\_histogram(binwidth = 0.25)

* **Unusual Values**

>ggplot(diabetes) + geom\_histogram(mapping = aes(x =Age), binwidth = 0.5)

* Missing Values

>ggplot(data = diabetes, mapping = aes(x = x, y = y)) + geom\_point()

* **A Categorical and Continuous Variable**

> ggplot(data = diabetes, mapping = aes(x = Glucose)) + geom\_freqpoly(mapping = aes(color = "Outcome"), binwidth = 500)

>ggplot(diabetes) + geom\_bar(mapping = aes(x = Glucose))

>ggplot(data = diabetes, mapping = aes(x = Outcome, y = Glucose)) + geom\_boxplot()

* **Two Categorical Variables**

ggplot(data = diabetes) + geom\_count(mapping = aes(x = BMI, y = Insulin))

* **Two Continuous Variables**

>ggplot(data = diabetes) + geom\_point(mapping = aes(x = Insulin, y = BMI))

ggplot(data = diabetes) + geom\_point(

+ mapping = aes(x = Age, y = BMI),

+ alpha = 10 / 100

+ )

>ggplot(data = diabetes, mapping = aes(x = Age, y = BMI)) + geom\_boxplot(mapping = aes(group = cut\_width(Outcome, 0.1)))

>ggplot(data = diabetes, mapping = aes(x = Age, y = BMI)) +

geom\_boxplot(mapping = aes(group = cut\_number(BloodPressure, 10)))

>ggplot(data = diabetes) + geom\_point(mapping = aes(x = BloodPressure, y = Age)) +

+ coord\_cartesian(xlim = c(4, 11), ylim = c(4, 11))

* **Patterns and Models**

>ggplot(data = faithful) + geom\_point(mapping = aes(x = eruptions, y = waiting))

> ggplot(data = diabetes) + geom\_boxplot(mapping = aes(x = Insulin, y = BloodPressure))

> ggplot(data = diabetes) + geom\_boxplot(mapping = aes(x = Insulin, y = BloodPressure))

* ggplot2 Calls

>ggplot(data = diabetes, mapping = aes(x = BloodPressure)) + geom\_freqpoly(binwidth = 0.25)

>ggplot(diabetes, aes(BMI)) + geom\_freqpoly(binwidth = 0.25)

>diabetes %>%+ count(Age, BMI) %>%+ ggplot(aes(BMI, Age, fill = n)) +geom\_tile()