**#DATA WRANGLING**

**-Filter Verb :**

**gapminder %>%** -> Datasets

**Filter(year==2007, country == “United States”)**

**-Sort :**

**Gapminder %>%**

**Arrange(gdpPercap)** -Ascend

**Gapminder %>%**

**Arrange(desc(gdpPercap))** -Descend

**-Mutate :**

**gapminder %>%**

**mutate(lifeExp = lifeExp\*12) -> Change existing column**

**gapminder %>%**

**mutate(lifeExpMonths = lifeExp\*12) -> Add new column**

**#Visualization**

**\*ggplot(dataset, aes(x, y)) +geom\_point()**

\***log\_scale :** **ggplot(dataset, aes(x, y)) +geom\_point() + scale\_x\_log10()**

**\*Adding Color : ggplot(gapminder\_2007, aes(pop, lifeExp, color = continent)) +**

**geom\_point()+ scale\_x\_log10() + scale\_y\_log10()**

**\*Adding Size : ggplot(gapminder\_2007, aes(pop, lifeExp, color = continent, size = gdpPercap)) + geom\_point()+ scale\_x\_log10() + scale\_y\_log10()**

**\*Faceting : ggplot(gapminder\_2007, aes(pop, lifeExp)) +**

**geom\_point()+ scale\_x\_log10() + scale\_y\_log10() +facet\_wrap (~continent)**

**#Grouping and Summarizing**

**-Summarize :**

**gapminder %>%**

**summarize(medianLifeExp = median(lifeExp))**

**-Double Summarizing :**

**gapminder %>%**

**filter(year == 1957) %>%**

**summarize(medianLifeExp = median(lifeExp),**

**maxGdpPercap = max(gdpPercap))**

**-Group\_By :**

**gapminder %>%**

**filter(year == 1957) %>%**

**group\_by( (continent)) %>%**

**summarize(medianLifeExp = median(lifeExp),**

**maxGdpPercap = max(gdpPercap))**

**-Double Group\_by :**

**gapminder %>%**

**filter(year == 1957) %>%**

**group\_by(continent, year) %>%**

**summarize(medianLifeExp = median(lifeExp),**

**maxGdpPercap = max(gdpPercap))**

**-Visulaizing Summarized Data :**

**by\_year <- gapminder %>%**

**group\_by(year) %>%**

**summarize(medianLifeExp = median(lifeExp),**

**madxGpPercap = max(gdpPercap))**

**ggplot(by\_year, aes(year, medianLifeExp)) +geom\_point()**

**#TYPE OF VISUALIZATION**

**-Line Plot :**

**library(gapminder)**

**library(dplyr)**

**library(ggplot2)**

**by\_year <- gapminder %>%**

**group\_by(year) %>%**

**summarize(medianGdpPercap = median(gdpPercap))**

**ggplot(by\_year, aes(year, medianGdpPercap)) +geom\_line() +**

**expand\_limits(y=0)**

**-Bar Column :**

**library(gapminder)**

**library(dplyr)**

**library(ggplot2)**

**by\_continent <- gapminder %>%**

**filter(year == 1952) %>%**

**group\_by(continent) %>%**

**summarize(medianGdpPercap = median(gdpPercap))**

**ggplot(by\_continent, aes(continent, medianGdpPercap)) +geom\_col()**

**-Histogram :**

**library(gapminder)**

**library(dplyr)**

**library(ggplot2)**

**gapminder\_1952 <- gapminder %>%**

**filter(year==1952) %>%**

**mutate(pop\_by\_mil = pop/1000000)**

**ggplot(gapminder\_1952, aes(pop\_by\_mil)) +geom\_histogram(binwidth = 50)**

**\*note : Jika sebaran data terlalu beragam, pakai log scale**

**-Boxplot :**

**library(gapminder)**

**library(dplyr)**

**library(ggplot2)**

**gapminder\_1952 <- gapminder %>%**

**filter(year==1952)**

**ggplot(gapminder\_1952, aes(continent, gdpPercap)) +geom\_boxplot() +scale\_y\_log10()**