Learning R – Tydiverse

# Load the gapminder package

library(gapminder)

# Load the dplyr package

library(dplyr)

# verbs: filter to subset, arrange to sort, mutate to change/create/add new #variables 🡪 dplyr operators

# Look at the gapminder dataset

Gapminder

#using actions:

#A pipe %>% means that you feed everything before the sign to the action #(verb) after it

Your\_Data %>%

filter(condition)

# Filter the gapminder dataset for the year 1957

gapminder %>%

filter(year == 1957)

# Filter for China in 2002

gapminder %>%

filter(coutry == "China", year == 2002)

# Sort in ascending order of lifeExp

gapminder %>%

arrange(lifeExp)

# Sort in descending order of lifeExp

gapminder %>%

arrange(desc(lifeExp))

# Filter for the year 1957, then arrange in descending order of population

gapminder %>%

filter(year == 1957) %>%

arrange(desc(pop))

# Use mutate to change lifeExp to be in months

gapminder %>%

mutate(lifeExp = lifeExp \* 12)

# Use mutate to create a new column called lifeExpMonths

gapminder %>%

mutate(lifeExpMonths = lifeExp \* 12)

# Filter, mutate, and arrange the gapminder dataset

gapminder %>%

filter(year == 2007) %>%

mutate(lifeExpMonths = 12 \* lifeExp) %>%

arrange(desc(lifeExpMonths))

# THE SUMMARIZE VERBS

# Summarize to find the median life expectancy

gapminder %>%

summarize(medianLifeExp = median(lifeExp))

# Filter for 1957 then summarize the median life expectancy

gapminder %>%

filter(year == 1957) %>%

summarize(medianLifeExp = median(lifeExp))

# Filter for 1957 then summarize the median life expectancy and the maximum GDP per capita

gapminder %>%

filter(year == 1957) %>%

summarize(medianLifeExp = median(lifeExp),

maxGdpPercap = max(gdpPercap))

#The group\_by verb

# Find median life expectancy and maximum GDP per capita in each year

gapminder %>%

group\_by(year) %>%

summarize(medianLifeExp = median(lifeExp),

maxGdpPercap = max(gdpPercap))

# Find median life expectancy and maximum GDP per capita in each continent in 1957

gapminder %>%

filter(year == 1957) %>%

group\_by(continent) %>%

summarize(medianLifeExp = median(lifeExp),

maxGdpPercap = max(gdpPercap))

# Find median life expectancy and maximum GDP per capita in each continent/year combination

gapminder %>%

group\_by(continent, year) %>%

summarize(medianLifeExp = median(lifeExp),

maxGdpPercap = max(gdpPercap))

CHAPTER II

Datvizualisation with ggplot2

# Load the ggplot2 package as well

library(gapminder)

library(dplyr)

library(ggplot2)

# some aesthetics that can be used: x, y, color, size

# Create gapminder\_1952

gapminder\_1952 <- gapminder %>%

filter(year == 1952)

# Change to put pop on the x-axis and gdpPercap on the y-axis

ggplot(gapminder\_1952, aes(x = pop, y = gdpPercap)) +

geom\_point()

# Create a scatter plot with pop on the x-axis and lifeExp on the y-axis

ggplot(gapminder\_1952, aes(x = pop, y = lifeExp)) +

geom\_point()

# Change this plot to put the x-axis on a log scale

ggplot(gapminder\_1952, aes(x = pop, y = lifeExp)) +

geom\_point() +

scale\_x\_log10()

ggplot(gapminder\_1952, aes(x = pop, y= gdpPercap)) +

geom\_point() +

scale\_x\_log10()+

scale\_y\_log10()

# add color and size to represent more variables in a scatter plot

# Scatter plot comparing pop and lifeExp, with color representing continent and size gdpPercap

# Add the size aesthetic to represent a country's gdpPercap

ggplot(gapminder\_1952, aes(x = pop, y = lifeExp, color = continent,

size = gdpPercap)) +

geom\_point() +

scale\_x\_log10()

# Faceting: divide plot into subplots to represent different categorical 3 # variables

# Scatter plot comparing pop and lifeExp, faceted by continent

ggplot(gapminder\_1952, aes(x = pop, y = lifeExp)) +

geom\_point() +

scale\_x\_log10() +

facet\_wrap(~continent)

# Scatter plot comparing gdpPercap and lifeExp, with color representing continent

# and size representing population, faceted by year

ggplot(gapminder, aes(x = gdpPercap, y = lifeExp, color = continent,

size = pop)) +

geom\_point()+

scale\_x\_log10() +

facet\_wrap(~year)

#visualizing summarized data

by\_year <- gapminder %>%

group\_by(year) %>%

summarize(medianLifeExp = median(lifeExp),

maxGdpPercap = max(gdpPercap))

# Create a scatter plot showing the change in medianLifeExp over time

ggplot(by\_year, aes(x = year, y = medianLifeExp)) +

geom\_point()+

expand\_limits(y = 0)

# Summarize medianGdpPercap within each continent within each year: by\_year\_continent

by\_year\_continent <- gapminder %>%

group\_by(continent, year) %>%

summarize(medianGdpPercap = median(gdpPercap))

# Plot the change in medianGdpPercap in each continent over time

ggplot(by\_year\_continent, aes(x = year, y = medianGdpPercap, color= continent)) +

geom\_point()+

expand\_limits(y = 0)

# Summarize the median GDP and median life expectancy per continent in 2007

by\_continent\_2007 <- gapminder %>%

filter(year == 2007) %>%

group\_by(continent) %>%

summarize(medianGdpPercap = median(gdpPercap), medianLifeExp = median(lifeExp))

# Use a scatter plot to compare the median GDP and median life expectancy

ggplot(by\_continent\_2007, aes(x = medianGdpPercap, y = medianLifeExp,

color = continent))+

geom\_point()

# Lineplots

# Summarize the median gdpPercap by year, then save it as by\_year

by\_year <- gapminder %>%

group\_by(year) %>%

summarize(medianGdpPercap = median(gdpPercap))

# Create a line plot showing the change in medianGdpPercap over time

ggplot(by\_year, aes(x = year, y = medianGdpPercap)) +

expand\_limits(y = 0) +

geom\_line()

# Summarize the median gdpPercap by year & continent, save as by\_year\_continent

by\_year\_continent <- gapminder %>%

group\_by(year, continent) %>%

summarize(medianGdpPercap = median(gdpPercap))

# Create a line plot showing the change in medianGdpPercap by continent over time

ggplot(by\_year\_continent, aes(x = year, y= medianGdpPercap, color = continent)) +

expand\_limits(y = 0) +

geom\_line()