R Coding

## Visualizing Data using ggplot

library(gapminder)

View(gapminder)

gapminder %>%

filter (continent %in% c("Africa", "Europe")) %>%

filter (gdpPercap < 30000) %>%

ggplot(aes(x = gdpPercap,

y = lifeExp,

size = pop,

color = year))+

geom\_point()+

facet\_wrap(~continent) +

labs (title = "Life expectancy explained by GDP",

x = " GDP per capita",

y = "Life expectancy")

library(tidyverse)

library(ggridges)

library(patchwork)

library(viridis)

library(hrbrthemes)

library(gapminder)

theme\_set(theme\_bw())

attach(starwars)

starwars %>%

select(name,height,mass,gender,hair\_color) %>%

head()

p1 <- starwars %>%

ggplot(aes(x = height)) +

geom\_histogram(binwidth =20,

show.legend = F,

alpha = 0.5) +

labs(title = "Histogram",

x = "Height",

y = "Count")

p2 <- starwars %>%

ggplot(aes(x = height)) +

geom\_density(aes(fill = "blue"),

show.legend = F,

alpha = 0.5) +

labs(title = "Density plot",

x = "Height",

y = "Probability")

p3 <- starwars %>%

ggplot(aes(x = height)) +

geom\_boxplot(show.legend = F,

alpha = 0.5) +

labs(title = "Boxplot",

x = "Height")

p3a <- starwars %>%

ggplot(aes(x = height,

y = 1)) +

geom\_violin(aes(fill = "blue"),

show.legend = F,

alpha = 0.5) +

labs(title = "Violin plot",

x = "Height")

(p1 / p2 | p3 / p3a) +

plot\_annotation(title = "Single numeric variable",

theme = theme(plot.title = element\_text(size = 18,

colour = "blue"))+

theme(text = element\_text("mono"))

p4 <- starwars %>%

drop\_na(eye\_color) %>%

filter(eye\_color %in% c("black", "blue", "yellow")) %>%

ggplot(aes(x = eye\_color)) +

geom\_bar(stat = "count", alpha = 0.5) +

labs(title = "Barplot",

x = "Eye color",

y = "Count")

p5 <- starwars %>%

drop\_na(eye\_color, gender) %>%

filter(eye\_color %in% c("black", "blue", "yellow")) %>%

ggplot(aes(eye\_color, fill = gender)) +

geom\_bar(stat = "count", alpha = 0.5,

show.legend = F) +

labs(title = "Stacked barplot",

x = "Eye color",

y = "Count")

p5a <- starwars %>%

drop\_na(eye\_color, gender) %>%

filter(eye\_color %in% c("black", "blue", "yellow")) %>%

ggplot(aes(eye\_color, fill = gender)) +

geom\_bar(stat = "count", alpha = 0.5,

position = "dodge",

show.legend = F) +

labs(title = "Grouped barplot",

x = "Eye color",

y = "Count")

p5b <- starwars %>%

drop\_na(eye\_color, gender) %>%

filter(eye\_color %in% c("black", "blue", "yellow")) %>%

ggplot(aes(eye\_color, fill = gender)) +

geom\_bar(stat = "count", alpha = 0.5,

position = "fill",

show.legend = T) +

labs(title = "Percentage barplot",

x = "Eye color",

y = "Count") +

theme(legend.position = "bottom")

((p4 | p5) / (p5a | p5b))+

plot\_annotation(title = "One or more categorical variable",

theme = theme(plot.title = element\_text(size = 18,

colour = "blue"))) +

theme(text = element\_text("mono"))

#Lollipop Data Viz

view(chickwts)

chickwts %>%

group\_by(feed) %>%

mutate(mean\_by\_feed = mean(weight)) %>%

ungroup() %>%

mutate(feed = fct\_reorder(feed, mean\_by\_feed)) %>%

ggplot(aes(feed, weight, colour = feed,

show.legend = F)) +

coord\_flip() +

geom\_jitter(show.legend = F,

size = 4,

alpha = 0.2,

width = 0.05) +

stat\_summary(fun = mean, geom = "point", size = 8, show.legend = F) +

geom\_hline(aes(yintercept = mean(weight)),

colour = "gray70",

size = 0.9) +

geom\_segment(aes(x = feed, xend = feed,

y = mean(weight), yend = mean\_by\_feed),

size = 2, show.legend = F) +

labs(title = "Weight of chickens by feed group",

x = "Feed",y

= "Weight of chickens") +

theme(legend.position = "none") +

theme\_bw()

##Using Ridges

ggplot(lincoln\_weather, aes(x =`Mean Temperature [F]`, y =`Month`, fill = ..x..)) +

geom\_density\_ridges\_gradient(scale = 3, rel\_min\_height = 0.01,

alpha = 5) +

scale\_fill\_viridis(name = "Temp. [F]", option = "C") +

labs(title ='Temperatures in Lincoln NE in 2016') +

theme\_bw() +

theme(

legend.position="none",

panel.spacing = unit(0.1, "lines"),

strip.text.x = element\_text(size = 8))

## Statistical test using R

library(gapminder)

library(tidyverse)

## T-Test

gapminder %>%

filter(continent=="Africa") %>%

select(lifeExp) %>%

t.test(mu=50)

ttest <- gapminder %>%

filter(continent=="Africa") %>%

select(lifeExp) %>%

t.test(mu=50)

attributes(ttest)

ttest$p.value

## Two Sided test for difference of means

gapminder %>%

filter(continent %in% c("Africa", "Europe")) %>%

t.test(lifeExp~ continent, data = .,

alternative = "two.sided")

## One sided test for difference means

gapminder %>%

filter(country %in% c("Bangladesh", "Canada")) %>%

t.test(lifeExp ~ country, data = .,

alternative = "less",

conf.level = 0.95)

##Paired t-test

gapminder %>%

filter(year %in% c(1957,2007) &

continent =="Asia") %>%

mutate(year = factor(year, levels = c(2007,1957))) %>%

t.test(lifeExp ~ year, data = .,

paired = T)