Visualizations Practice 2

Abhilash Roy

3/25/2021

Including Plots

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

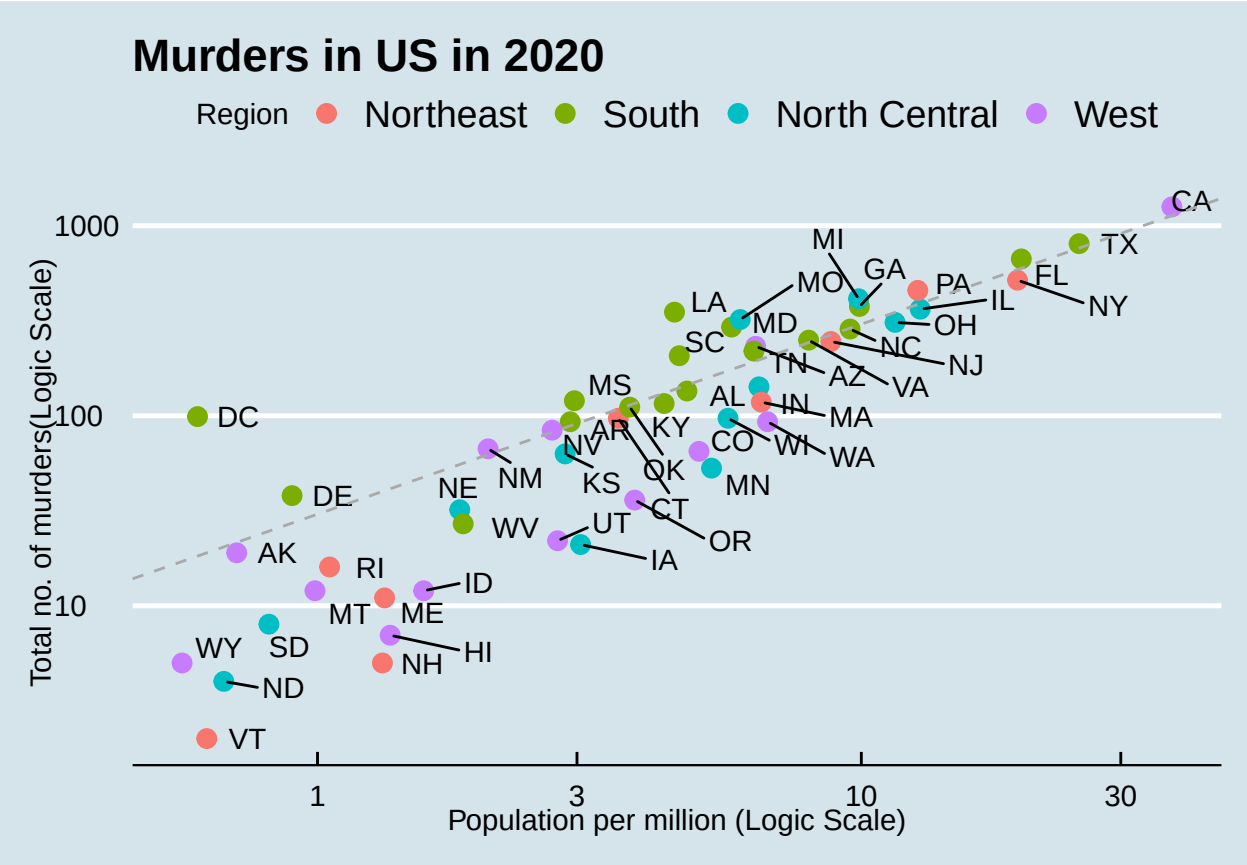
```
library(ggthemes)
library(ggrepel)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library(ggExtra)
library(cowplot)
```

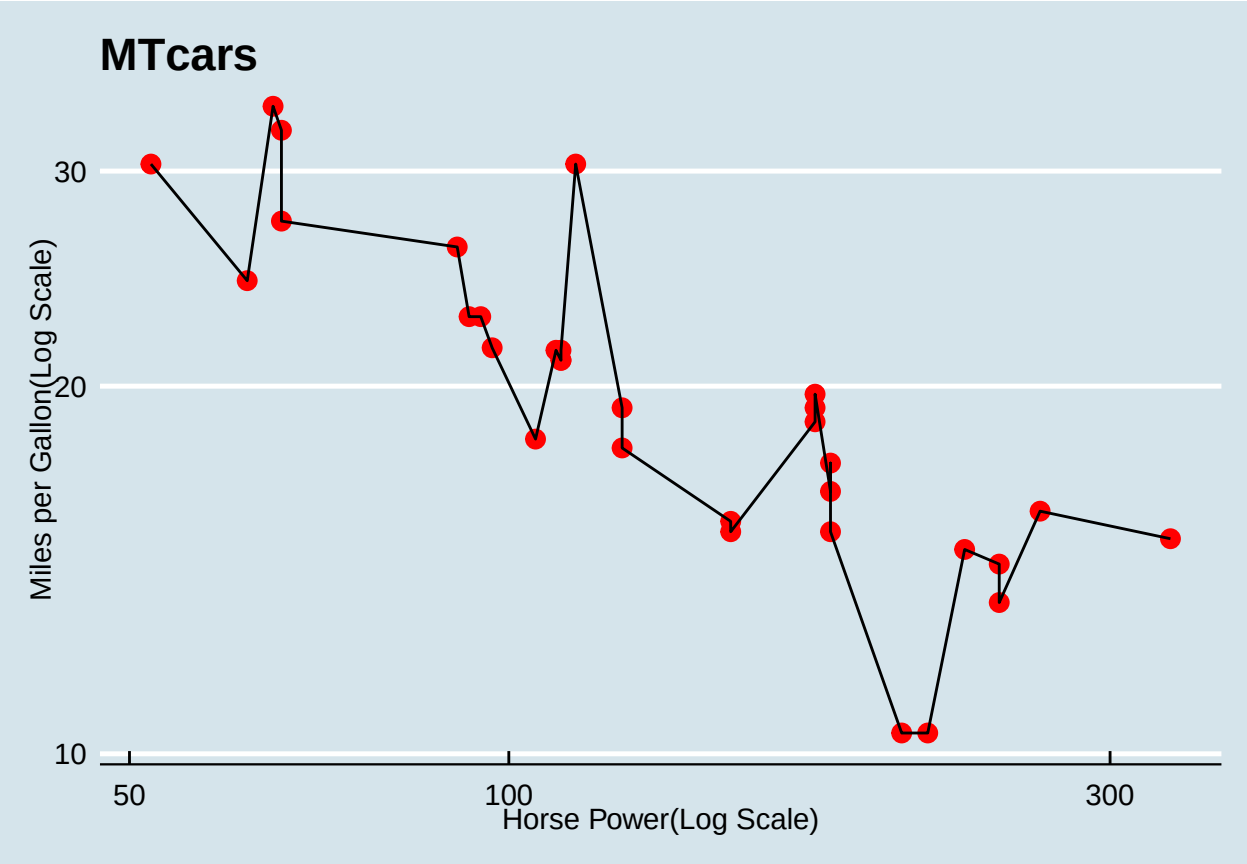
```
##
## Attaching package: 'cowplot'

## The following object is masked from 'package:ggthemes':
##
##     theme_map
```

```
library(gtools)
library(knitr)
library(rmarkdown)
library(tinytex)
#datasets
library(dslabs)
library(nycflights13)
library(NHANES)
library(titanic)

murder_rate <- murders %>% summarize(rate = sum(total)/sum(population) * 10^6) %>% pull(rate)

murders %>%
  ggplot(aes(population/10^6, total, label = abb)) +
  geom_point(aes(color = region), size = 3) + geom_text_repel(nudge_x = 0.075) +
  theme_economist() + scale_x_log10() + scale_y_log10() +
  xlab("Population per million (Logic Scale)") +
  ylab("Total no. of murders(Logic Scale)") +
  geom_abline(intercept = log10(murder_rate), lty = 2, color = "darkgrey") +
  ggtitle("Murders in US in 2020") + scale_color_discrete(name = "Region")
```

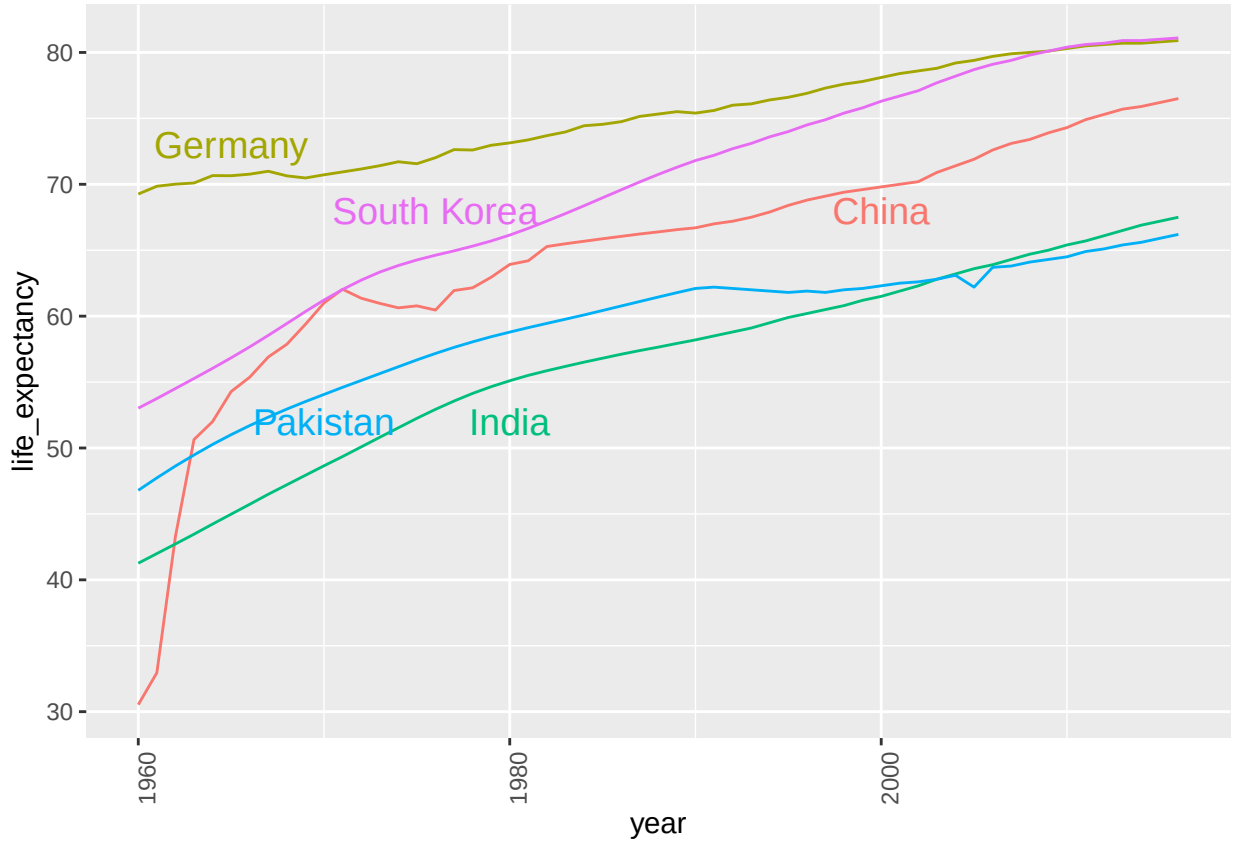



```
data(mtcars)

mtcars %>%
  ggplot(aes(hp, mpg)) + geom_point(size = 3, color = "red") +
```

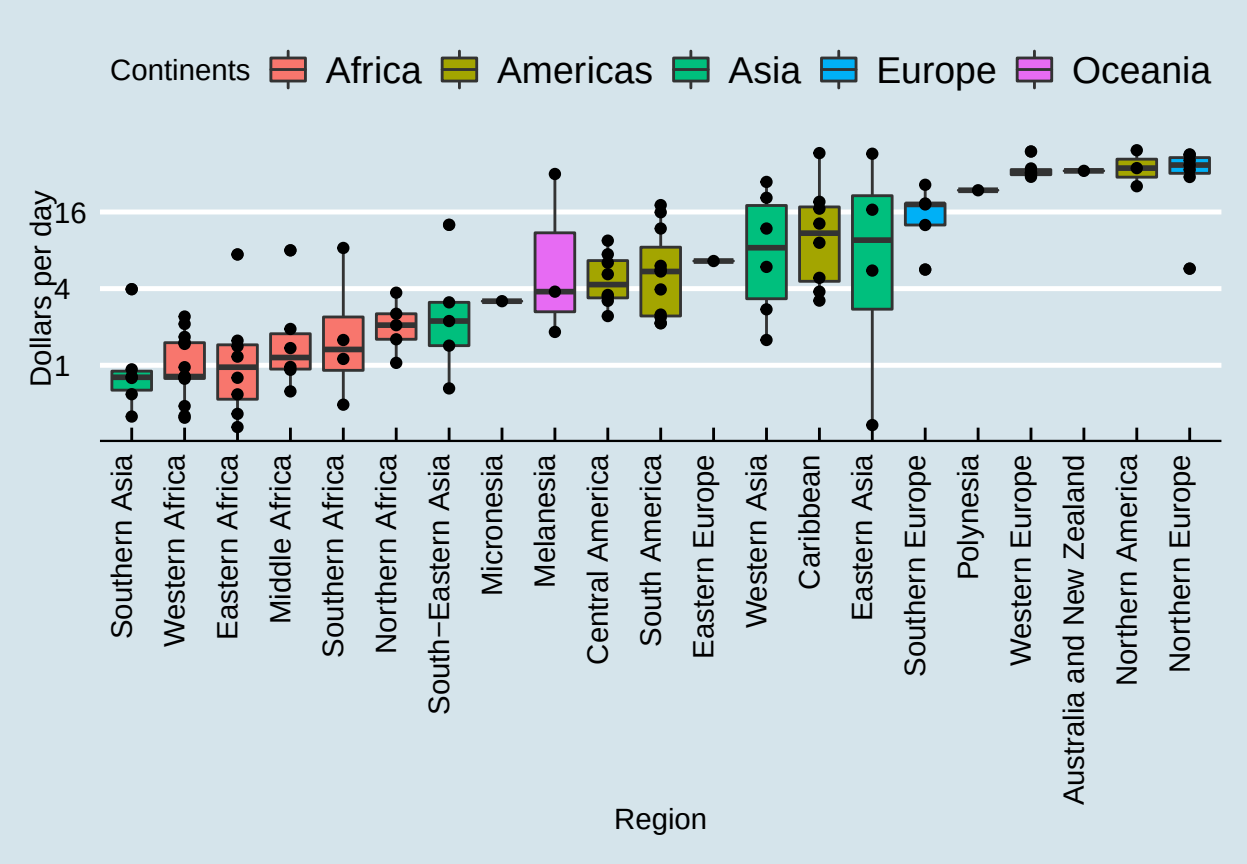
```
  xlab("Horse Power(Log Scale)") + ylab("Miles per Gallon(Log Scale)") +
  ggtitle("MTcars") + scale_x_log10() + scale_y_log10() +
  theme_economist() + geom_line()
```




```
labels_country <-
  data.frame(country = c("South Korea","Germany", "India", "China", "Pakistan"),
                     x = c(1976,1965,1980, 2000, 1970), y = c(68,73,52,68, 52))

gapminder %>%
  filter(country %in% c("South Korea","Germany","India","China","Pakistan")) %>%
  ggplot(aes(year, life_expectancy, col = country)) + geom_line() +
  geom_text(data = labels_country, aes(x, y, label = country), size = 5) +
  theme(legend.position = "none") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



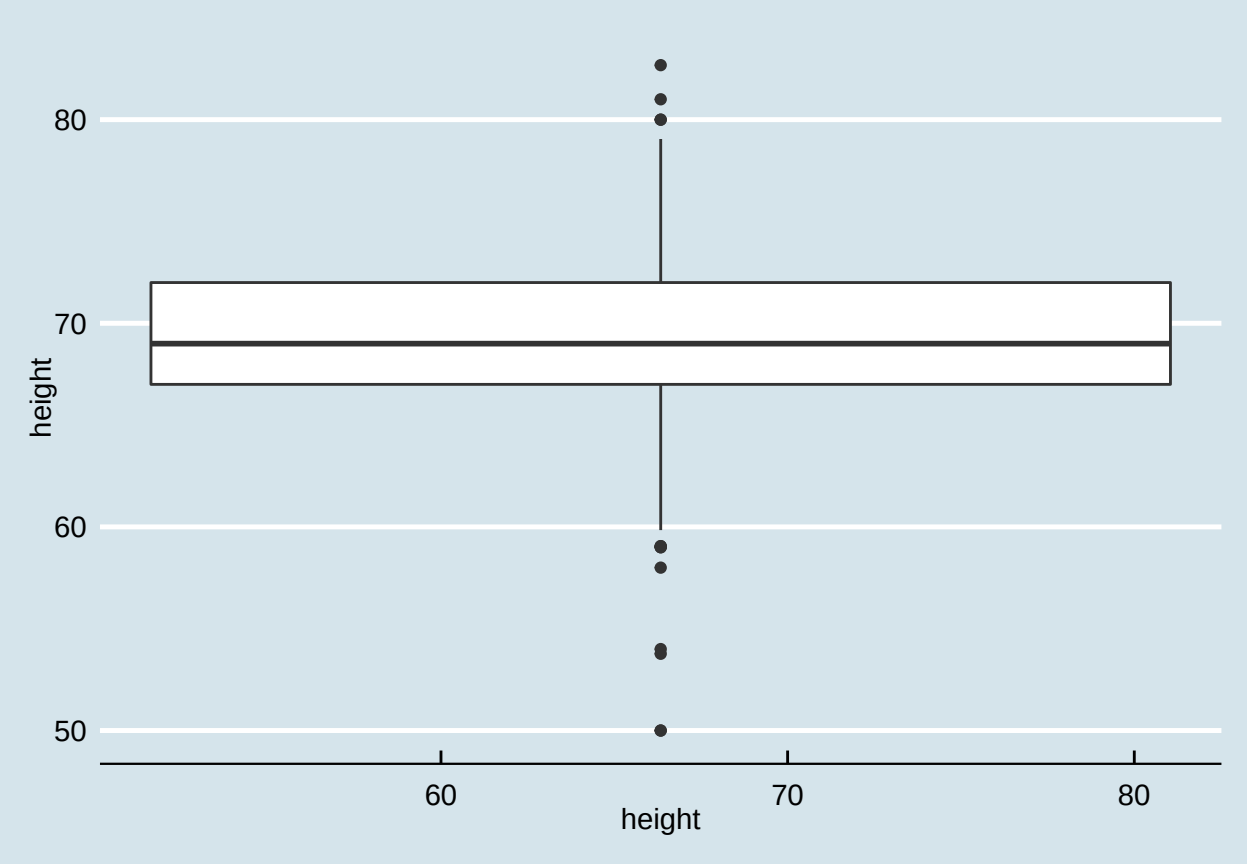

```
gapminder <- gapminder %>%
  mutate(dollars_per_day = gdp/population/365)
past_year <- 1970
gapminder %>%
  filter(year == past_year & !is.na(gdp)) %>%
  mutate(region = reorder(region, dollars_per_day, FUN = median)) %>%
  ggplot(aes(region, dollars_per_day)) + geom_boxplot(aes(fill = continent)) +
  theme_economist() +theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  xlab("Region") + ylab("Dollars per day") + scale_fill_discrete("Continents") +
  scale_y_continuous(trans = "log2") + geom_point()
```




```
view(gapminder)
```

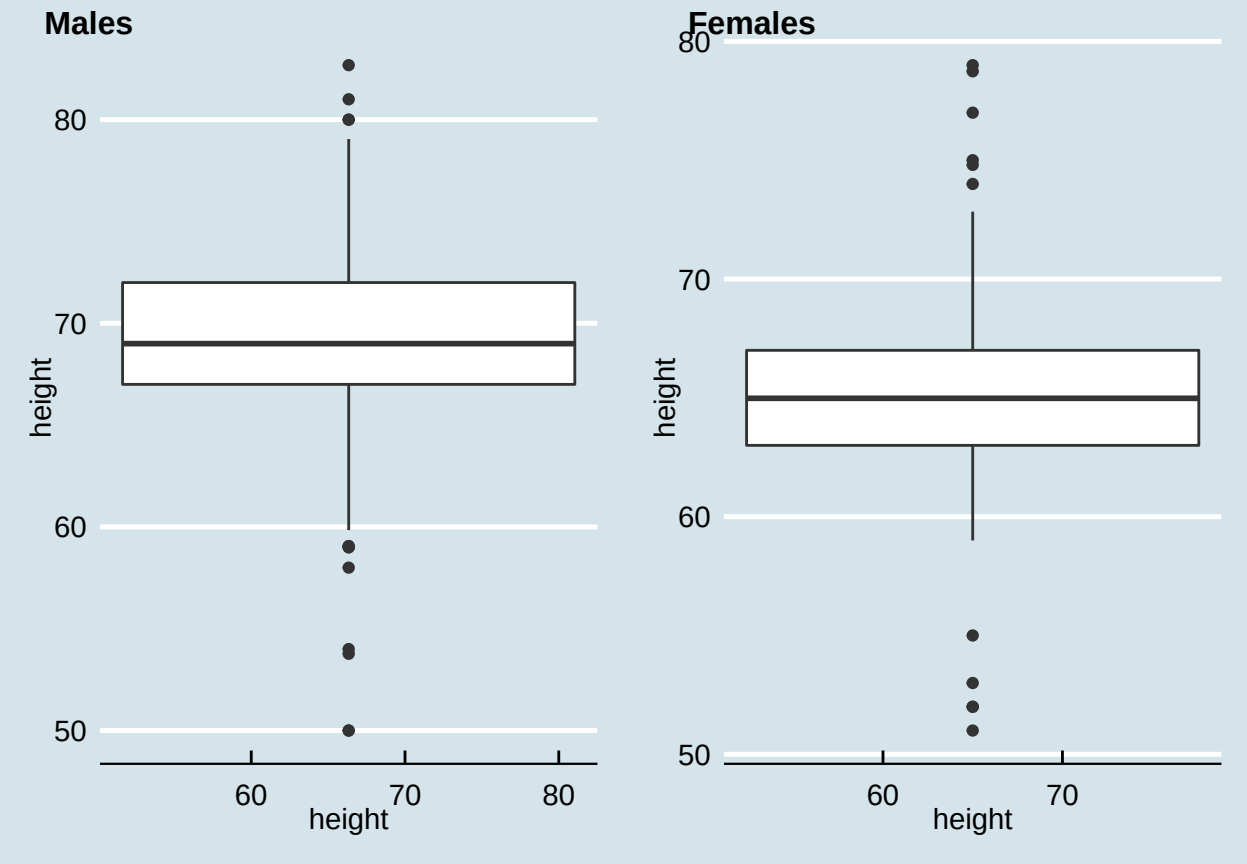
```
Males <-
  heights %>%
  filter(sex == "Male") %>% ggplot(aes(height,height)) +
  geom_boxplot() + theme_economist()

Males
```

```
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```

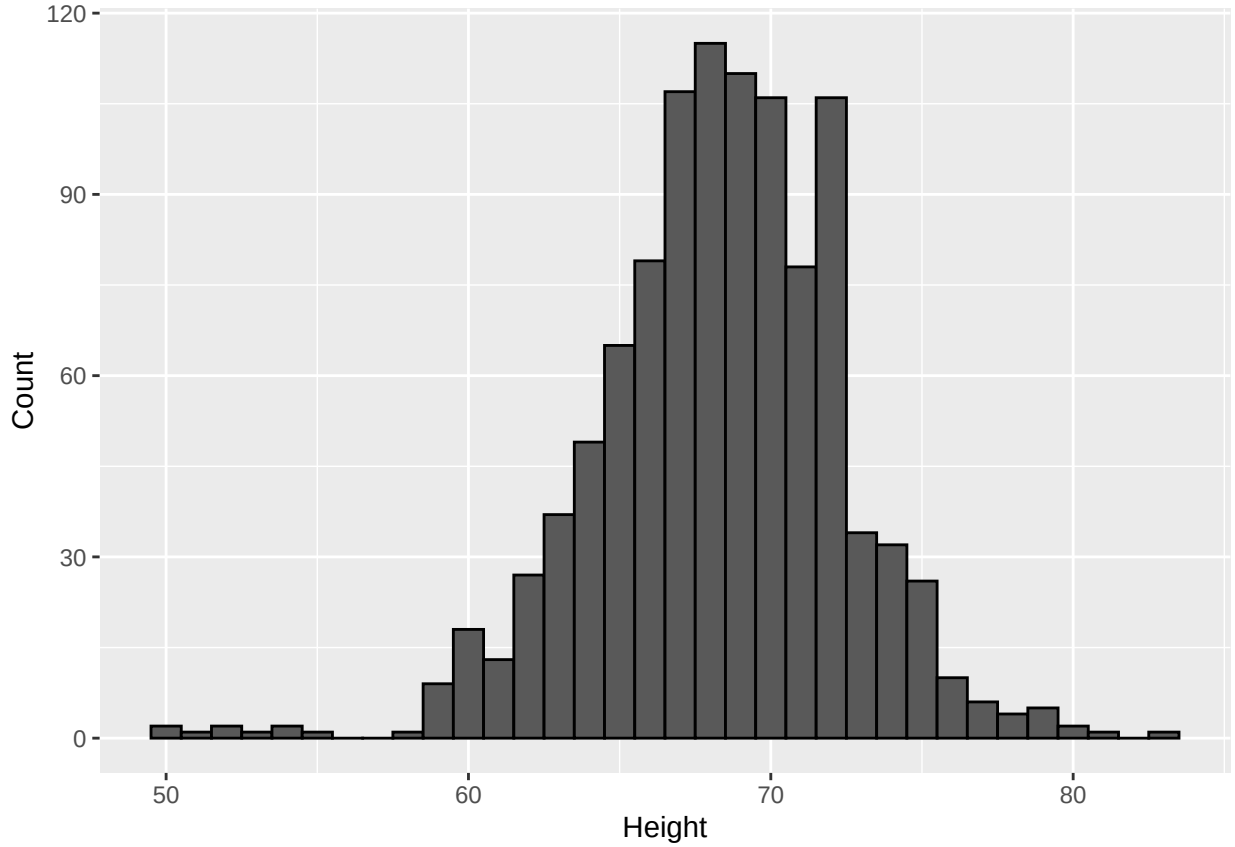



```
Females <-
  heights %>%
      filter(sex == "Female") %>% ggplot(aes(height,height)) +
      geom_boxplot() + theme_economist()


plot_grid(Males, Females, labels = c('Males', 'Females'), label_size = 12)
```

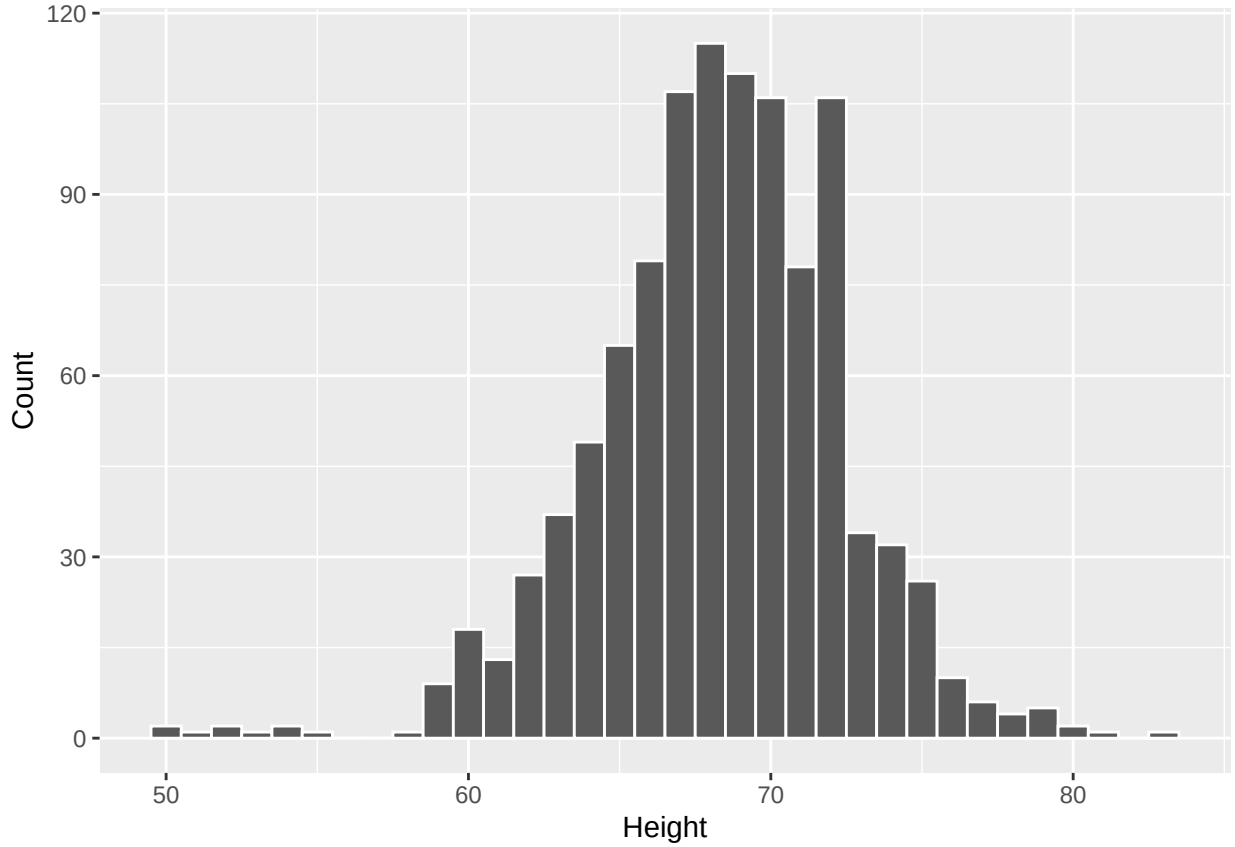
```
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?

## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```




```
heights %>%
  group_by(sex) %>% ggplot(aes(height)) +
  geom_histogram(binwidth = 1, color = "black") +
  xlab("Height") + ylab("Count")
```




```
heights %>%
  ggplot(aes(height)) +
  geom_histogram(binwidth = 1, color = "white") +
  xlab("Height") + ylab("Count")
```



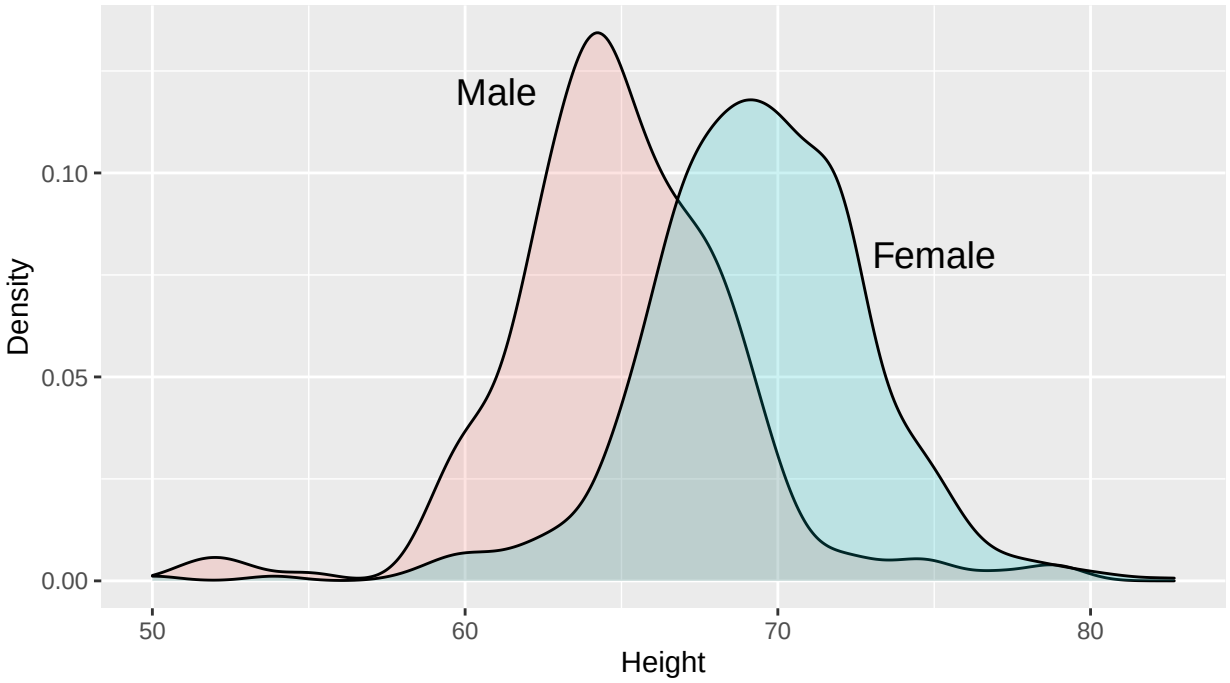

```
str(heights)
```

```
## 'data.frame':    1050 obs. of  2 variables:
##  $ sex   : Factor w/ 2 levels "Female","Male": 2 2 2 2 2 1 1 1 1 2 ...
##  $ height: num  75 70 68 74 61 65 66 62 66 67 ...
```

```
gender <- data.frame(sex = c("Male", "Female"), x = c(61,75), y = c(0.12, 0.08))
heights %>%
  ggplot(aes(height, fill = sex)) + geom_density(alpha = 0.2) +
  theme(legend.position = "none") +
  geom_text(data = gender, aes(x,y, label = sex), size = 5) +
  xlab("Height") + ylab("Density") +
  ggtitle(label = "Height across genders", subtitle = "(Density graph)") +
  theme(plot.title = element_text(color = "Blue", size = 20,
                                  face = "bold", hjust = 0.5)) +
  theme(plot.subtitle = element_text(color = "red", size = 15, hjust = 0.5)) +
  labs(caption = "By - Abhilash Roy") +
  theme(plot.caption = element_text(color = "black", size = 12))
```

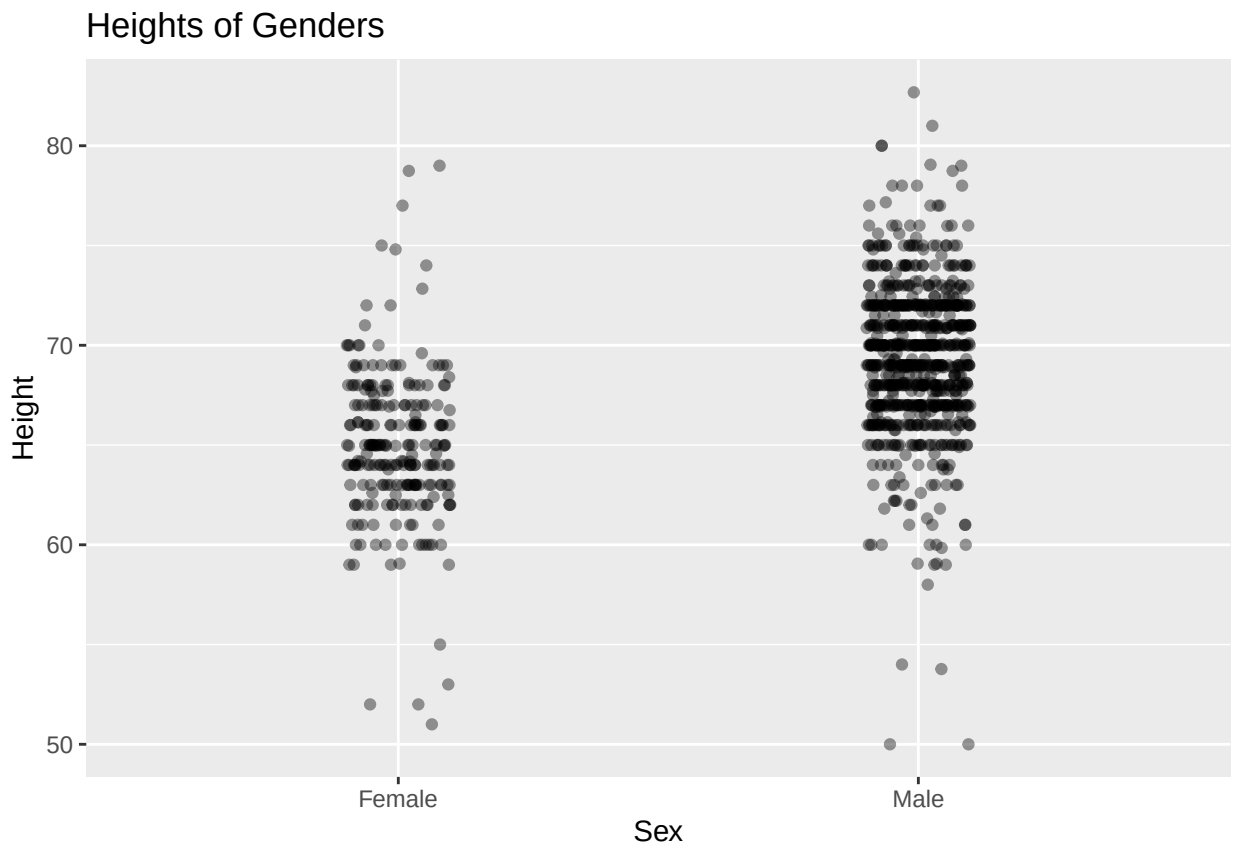
Height across genders

(Density graph)

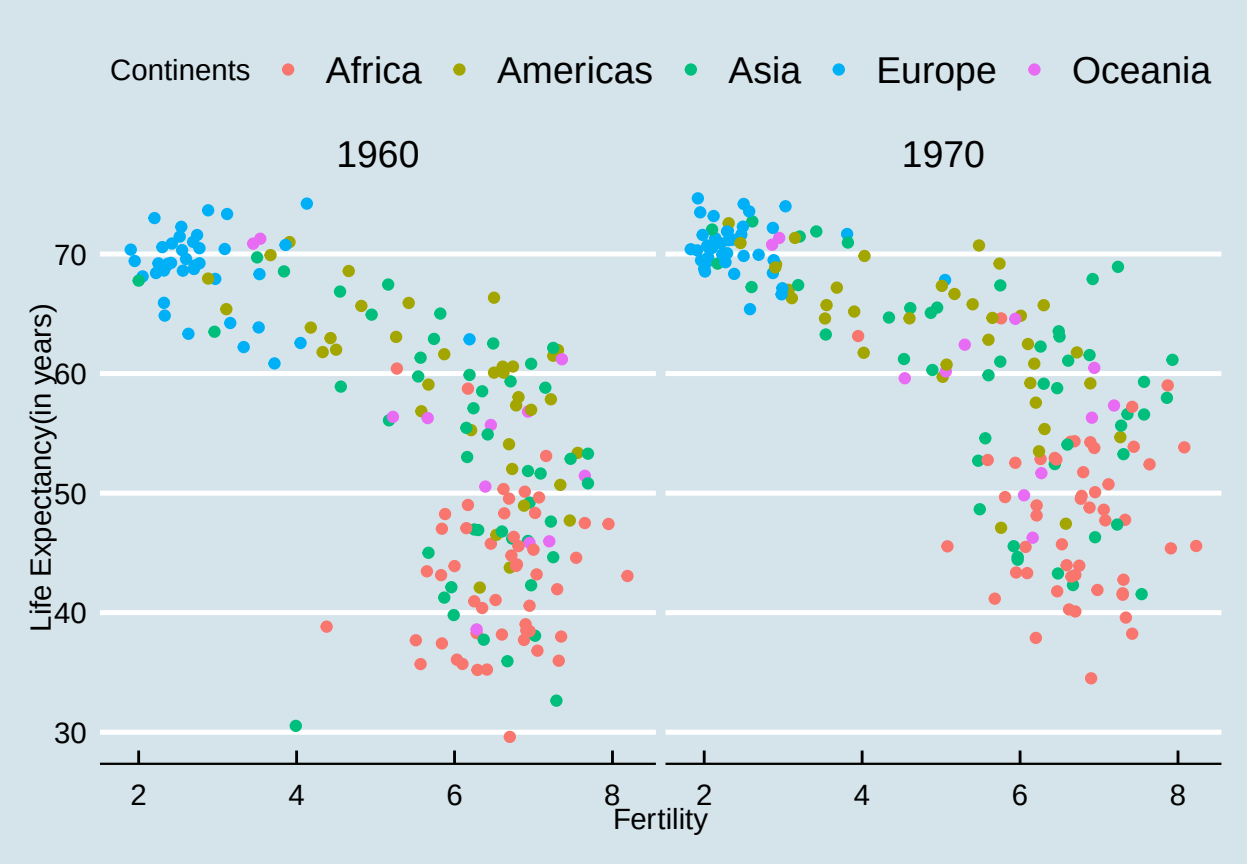


By − Abhilash Roy

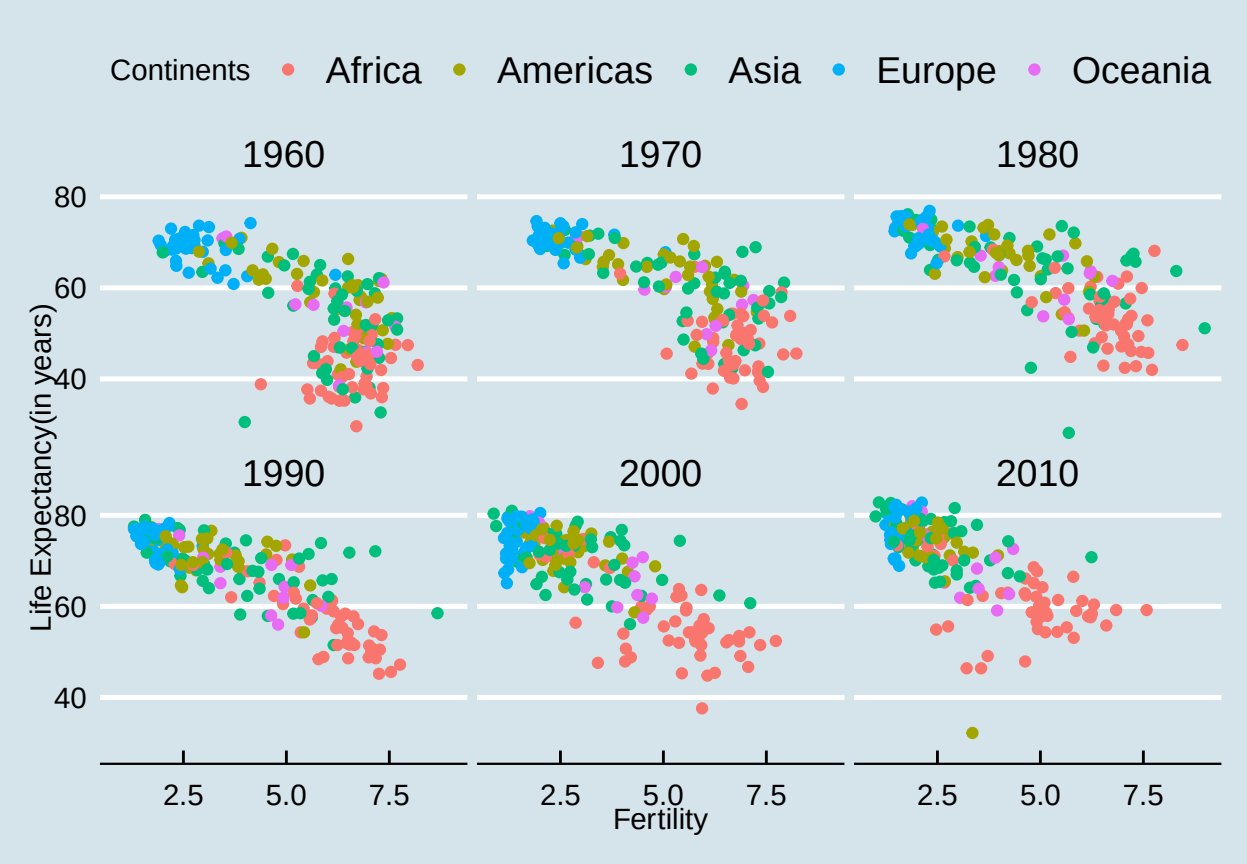
```
heights %>%
  ggplot(aes(sex, height)) + geom_jitter(width = 0.1, alpha = 0.4) +
  labs(x = "Sex", y = "Height", title = "Heights of Genders")
```



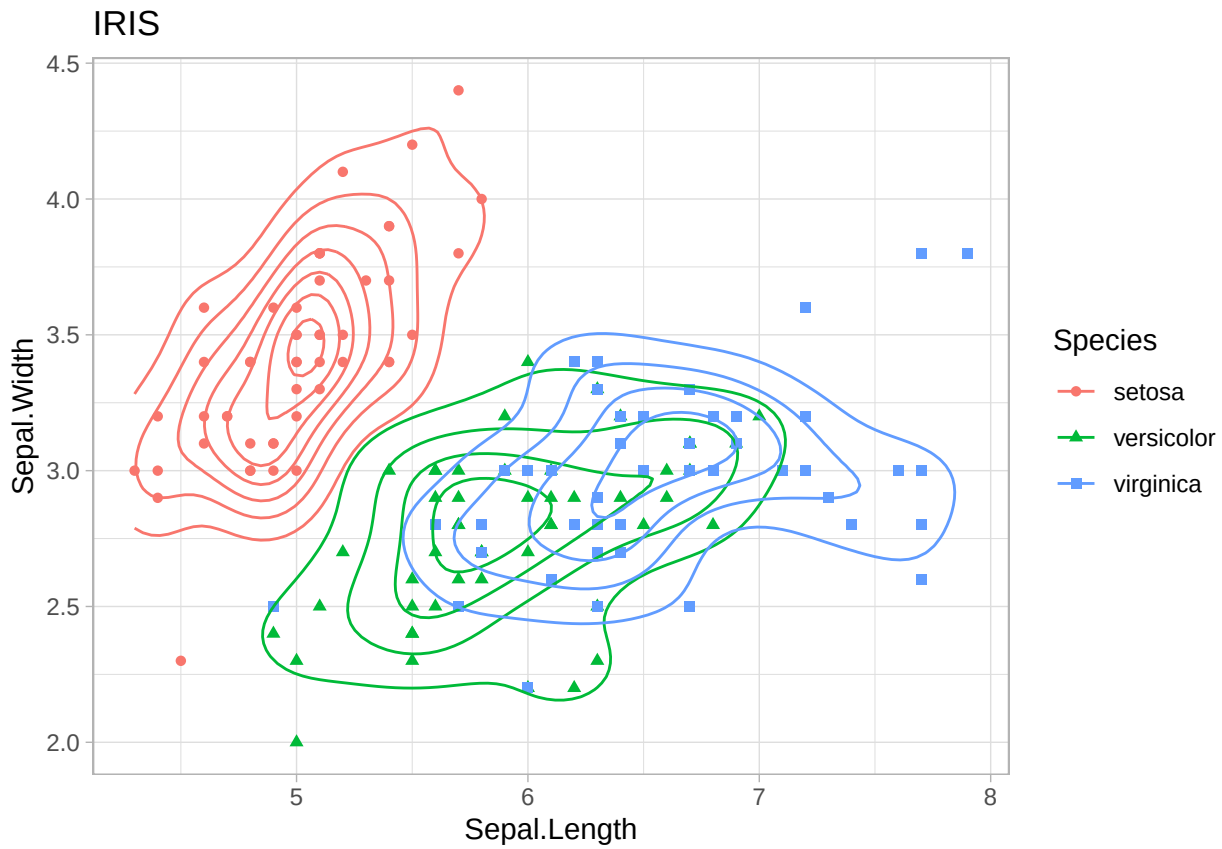

```
gapminder %>%
  filter(year %in% c(1960,1970)) %>% ggplot(aes(fertility,
  life_expectancy, color = continent)) + geom_point() + facet_grid(~year) +
  theme_economist() + xlab("Fertility") + ylab("Life Expectancy(in years)") +
  scale_color_discrete("Continents")
```




```
gapminder %>%
  filter(year %in% c(1960,1970,1980,1990,2000,2010)) %>%
  ggplot(aes(fertility, life_expectancy, color = continent)) + geom_point() +
  facet_wrap(~year) + theme_economist() + xlab("Fertility") +
  ylab("Life Expectancy(in years)") +
  scale_color_discrete("Continents")
```

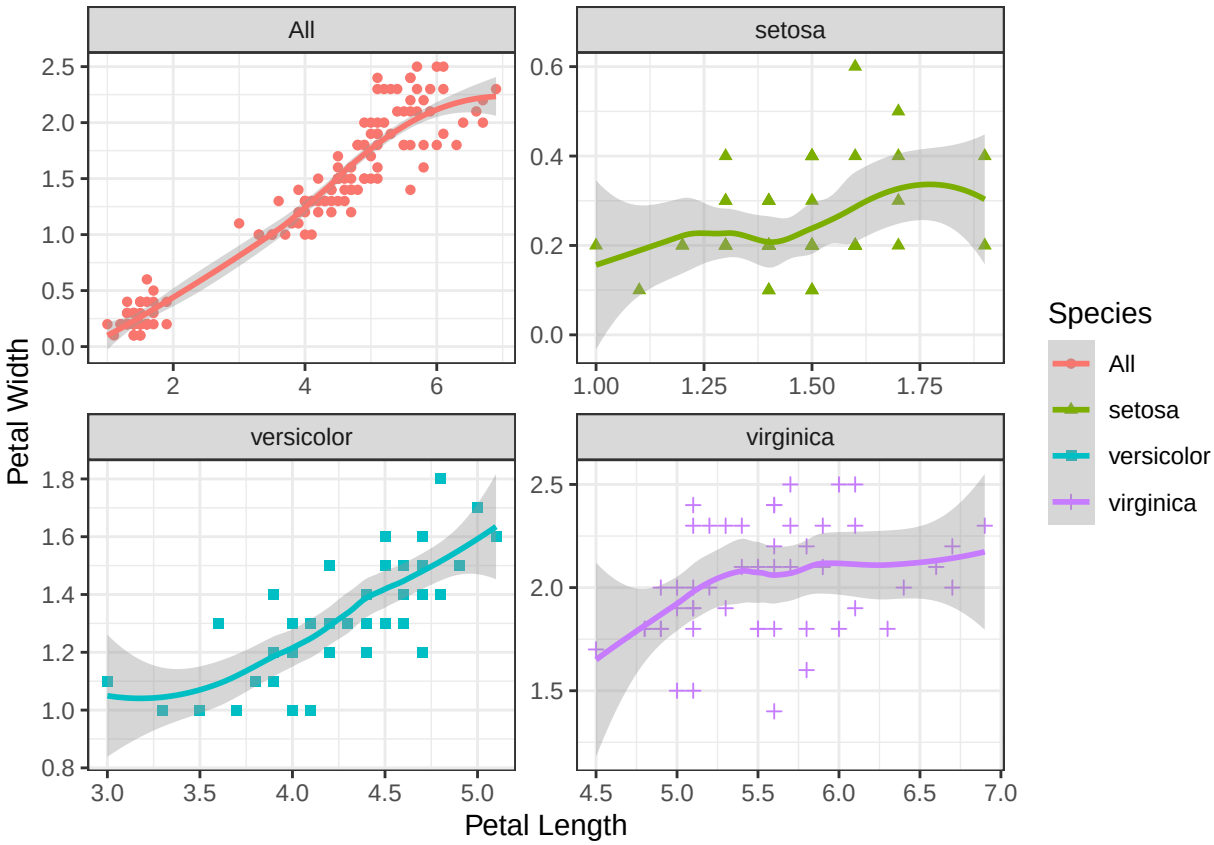



```
data(iris)

iris %>%
  ggplot(aes(Sepal.Length, Sepal.Width, color = Species, shape = Species)) +
  geom_point() +theme_light() + geom_density2d() + ggtitle("IRIS")
```

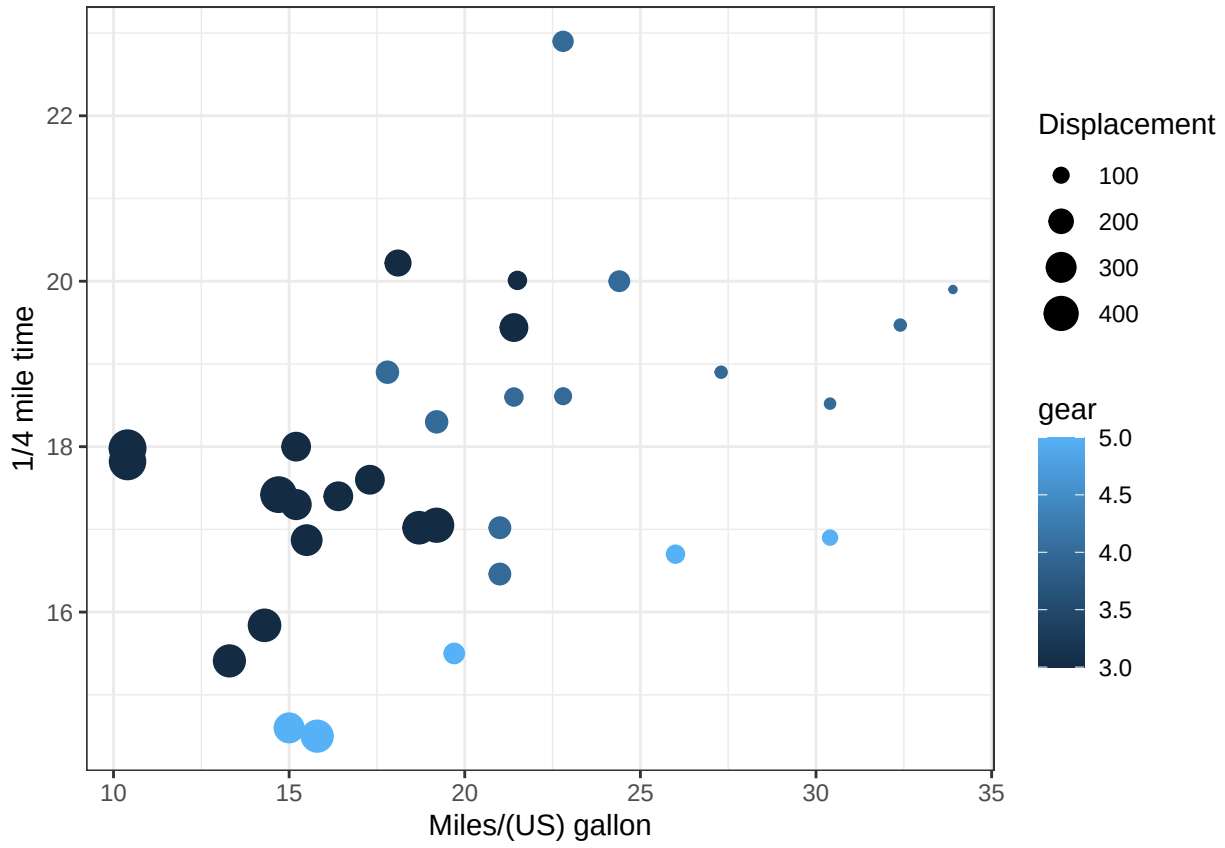



```
iris %>%
  mutate(Species = "All") %>% rbind(iris) %>%
  ggplot(aes(Petal.Length, Petal.Width, color = Species, shape =
  Species)) + geom_point() + theme_bw() + facet_wrap(~Species, scales =
  "free") + geom_smooth() + xlab("Petal Length") +ylab("Petal Width")
```

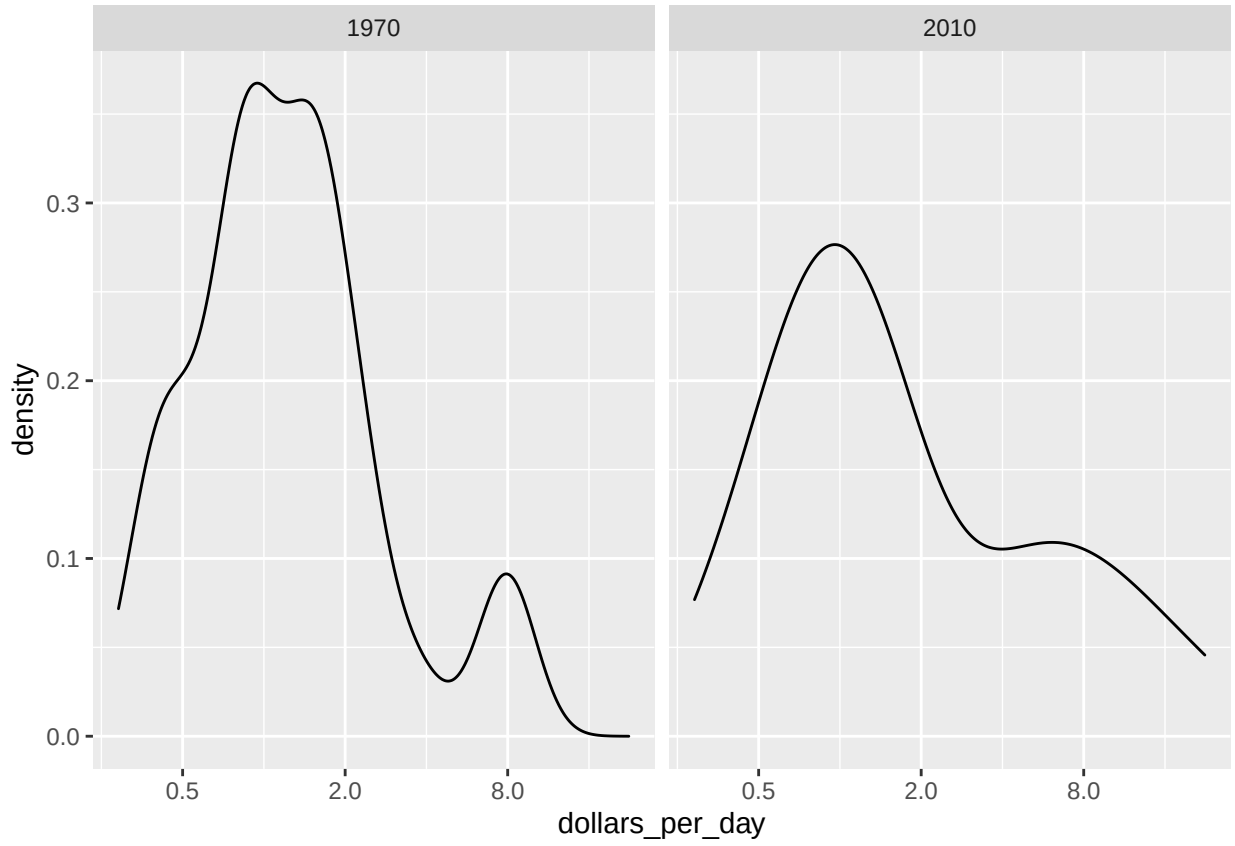
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



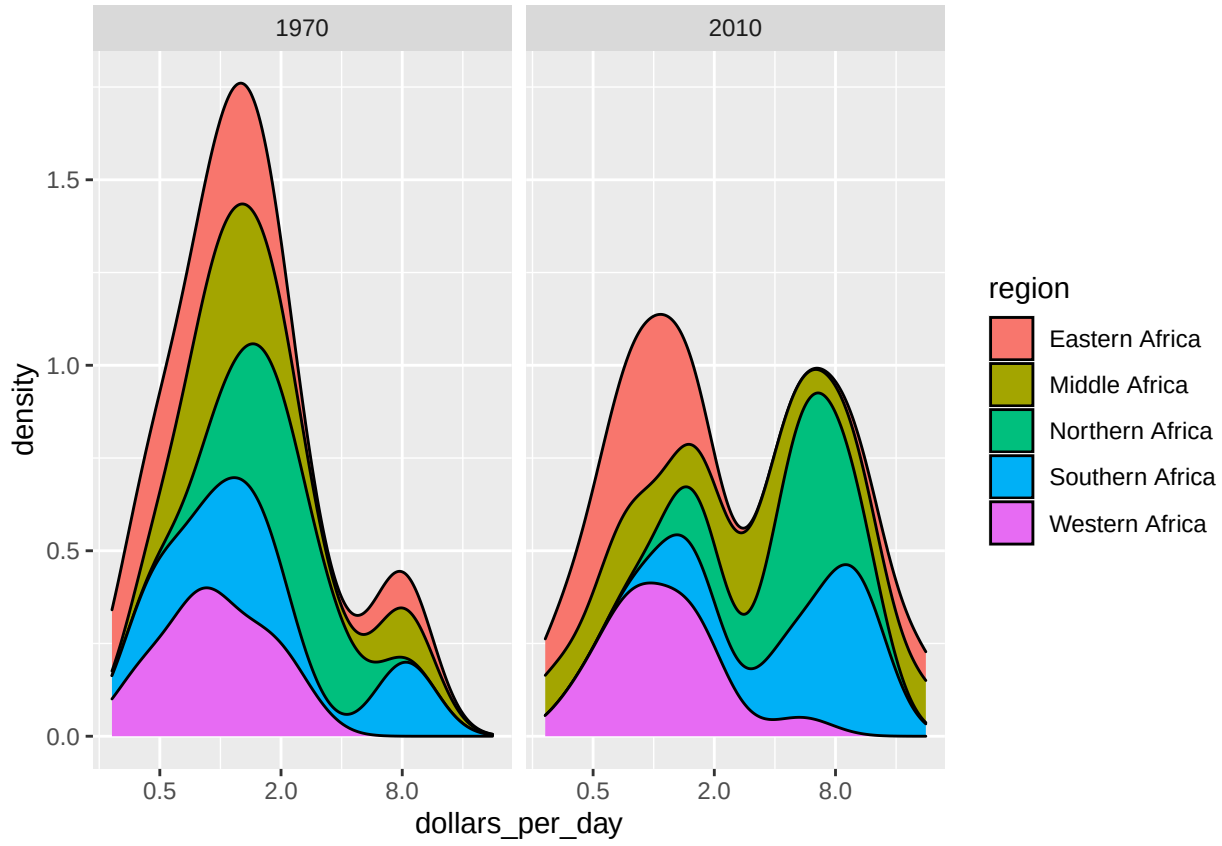

```
mtcars %>% ggplot(aes(mpg, qsec, color = gear, size = disp)) +
  geom_point() + xlab("Miles/(US) gallon") + ylab("1/4 mile time") +
  theme_bw() + scale_size_continuous("Displacement")
```



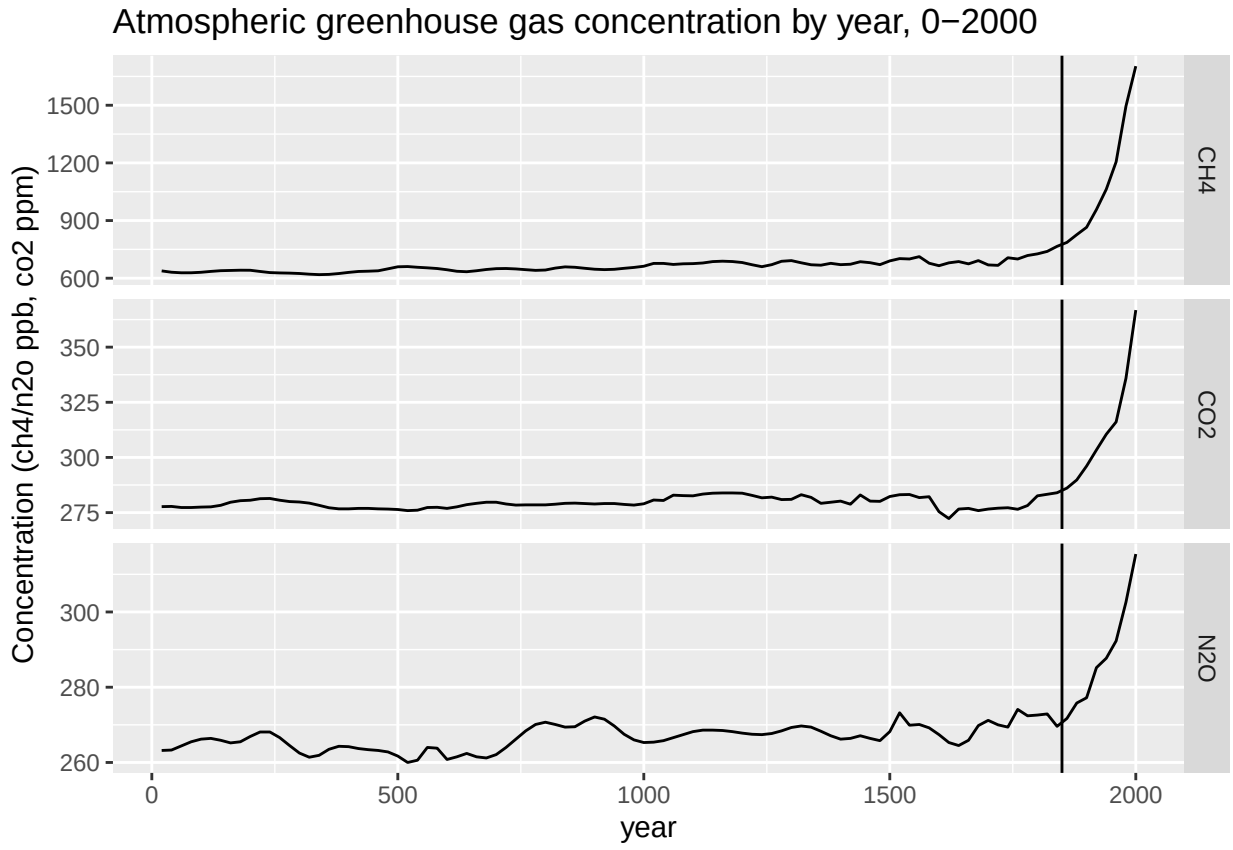

```
gapminder %>%
  mutate(dollars_per_day = gdp/population/365) %>%
  filter(continent == "Africa", year %in%
  c(1970,2010), !is.na(dollars_per_day)) %>%
  ggplot(aes(dollars_per_day)) + geom_density() + facet_grid(~year) +
  scale_x_continuous(trans = "log2")
```




```
gapminder %>%
  mutate(dollars_per_day = gdp/population/365) %>%
  filter(continent == "Africa", year %in% c(1970,2010),
  !is.na(dollars_per_day)) %>% ggplot(aes(dollars_per_day, fill = region)) +
  geom_density(bw= 0.5,position = "stack") + facet_grid(~year) +
  scale_x_continuous(trans = "log2")
```



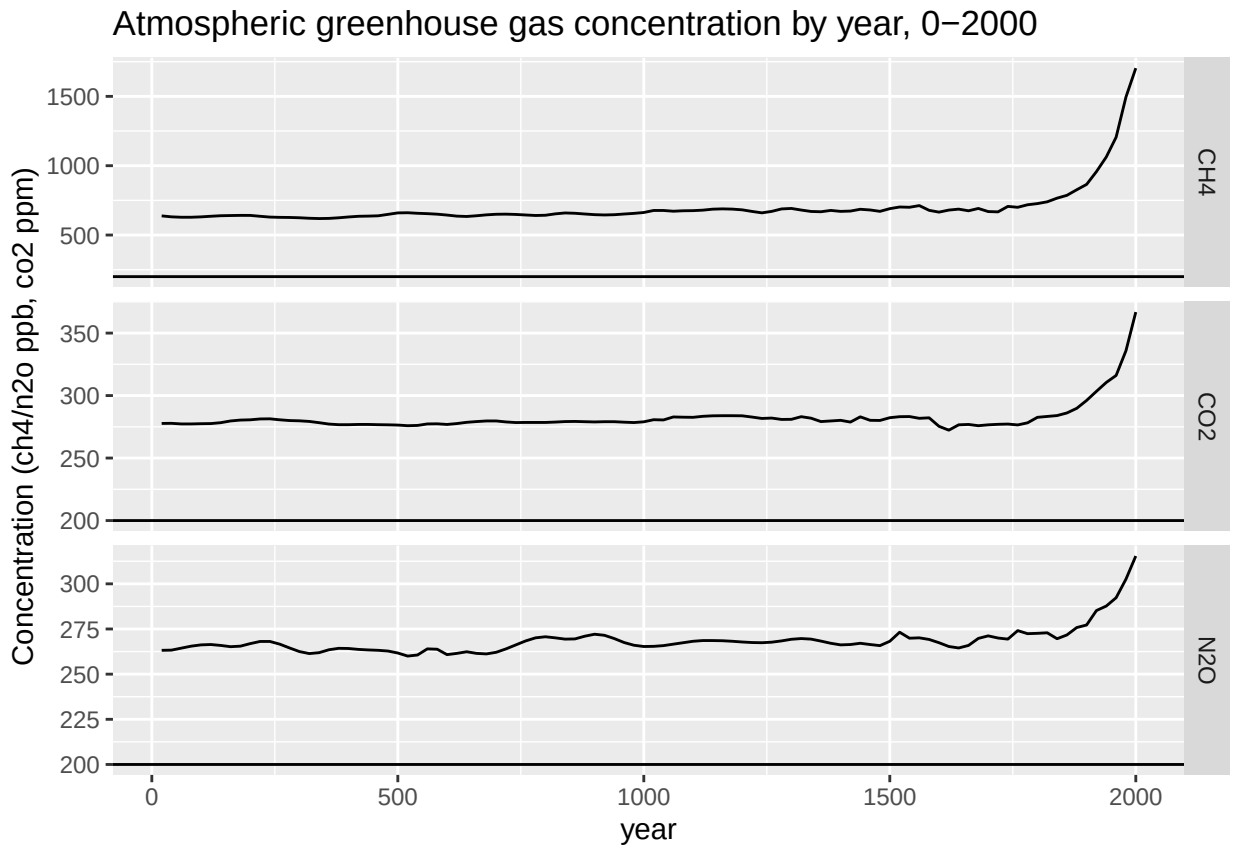

```
data(temp_carbon)
data(greenhouse_gases)
data(historic_co2)


greenhouse_gases %>%
  ggplot(aes(year,concentration)) +
  geom_line() +
  facet_grid(gas~., scales = "free") +
  ylab("Concentration (ch4/n2o ppb, co2 ppm)") +
  geom_vline(aes(xintercept = 1850)) +
  ggtitle("Atmospheric greenhouse gas concentration by year, 0-2000")
```




```
str(greenhouse_gases)
```

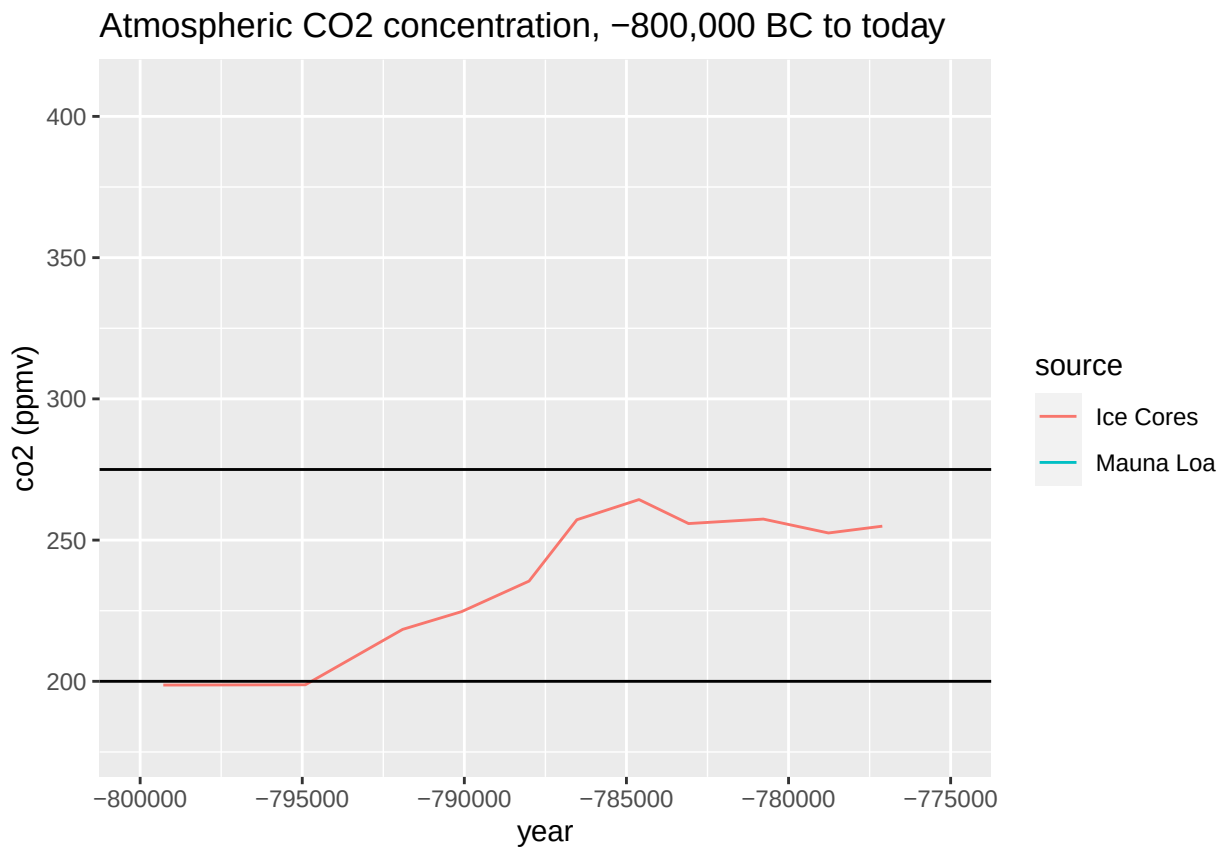
```
## 'data.frame':    300 obs. of  3 variables:
##  $ year         : num  20 40 60 80 100 120 140 160 180 200 ...
##  $ gas          : chr  "CO2" "CO2" "CO2" "CO2" ...
##  $ concentration: num  278 278 277 277 278 ...
```

```
greenhouse_gases %>%
  ggplot(aes(year,concentration)) +
  geom_line() +
  facet_grid(gas~., scales = "free") +
  ylab("Concentration (ch4/n2o ppb, co2 ppm)") +
  geom_hline(aes(yintercept = 200)) +
  ggtitle("Atmospheric greenhouse gas concentration by year, 0-2000")
```



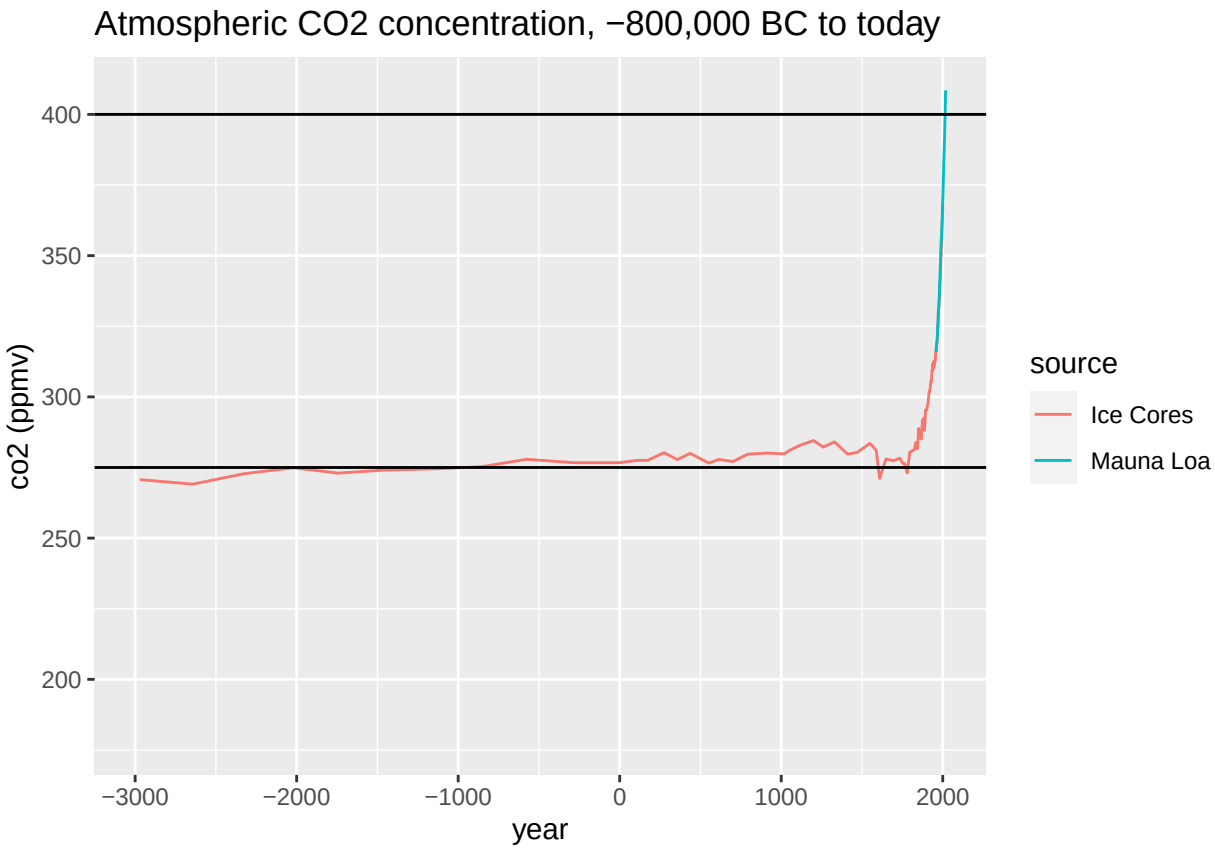

```
co2_time <- historic_co2 %>%
  ggplot(aes(year, co2, col = source)) + geom_line() +
  ggtitle("Atmospheric CO2 concentration, -800,000 BC to today") +
  ylab("co2 (ppmv)") + xlim(-800000, -775000) +
  geom_hline(aes(yintercept = 200)) +
  geom_hline(aes(yintercept = 275))
co2_time
```

```
## Warning: Removed 683 row(s) containing missing values (geom_path).
```



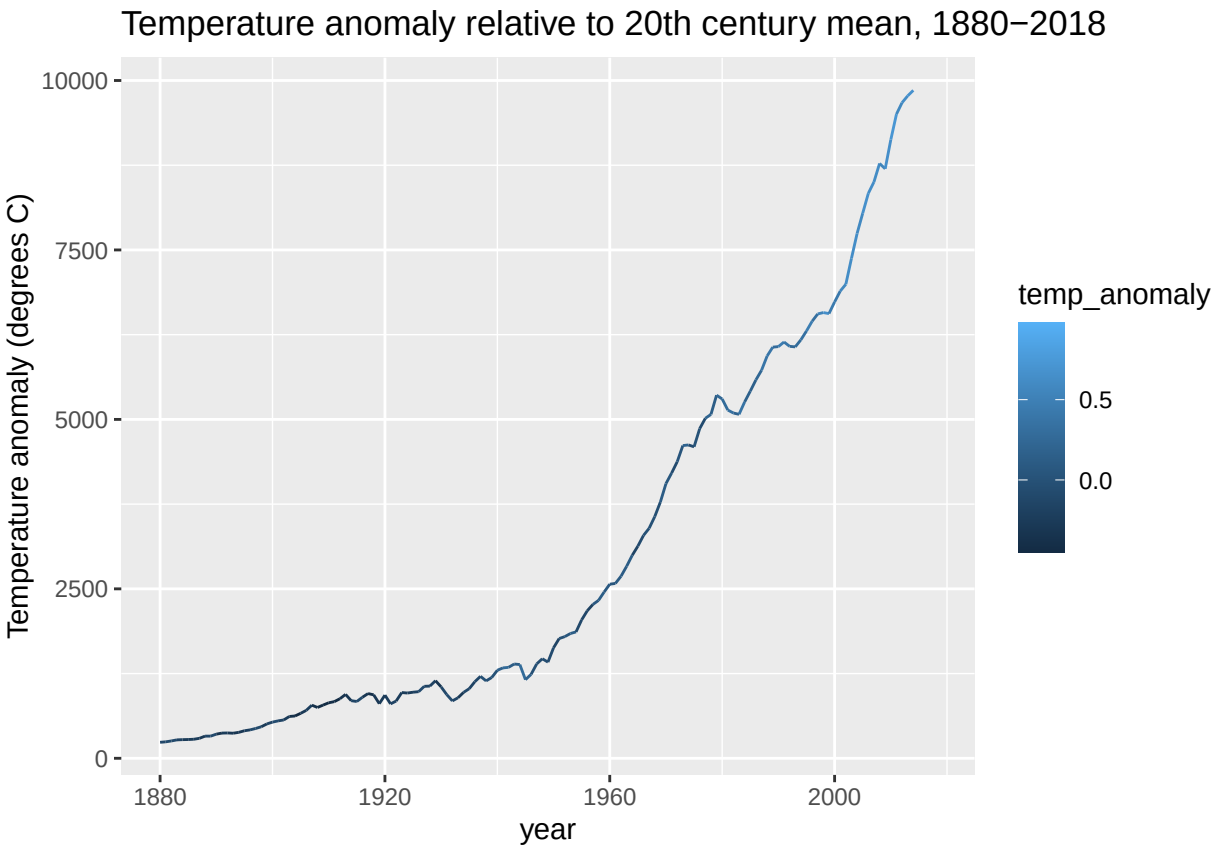

```
co2_time <- historic_co2 %>%
  ggplot(aes(year, co2, col = source)) + geom_line() +
  ggtitle("Atmospheric CO2 concentration, -800,000 BC to today") +
  ylab("co2 (ppmv)") + xlim(-3000, 2018) +
  geom_hline(aes(yintercept = 275)) +
  geom_hline(aes(yintercept = 400))
co2_time
```

```
## Warning: Removed 539 row(s) containing missing values (geom_path).
```




```
temp_carbon %>% filter(!is.na(temp_anomaly),!is.na(land_anomaly),
                       !is.na(ocean_anomaly)) %>%
  ggplot(aes(year, carbon_emissions, color = temp_anomaly)) +geom_line() +
  ylab("Temperature anomaly (degrees C)") +
  ggtitle("Temperature anomaly relative to 20th century mean, 1880-2018")
```

```
## Warning: Removed 4 row(s) containing missing values (geom_path).
```



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.