Week7-AE-RMarkdown

2024-03-27

```
library(tidyverse)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.0     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(scales)
```

```
##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
##
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
#install.packages('WDI')
library(WDI)
```

```
## Warning: package 'WDI' was built under R version 4.3.3
```

```
#install.packages('ggrepel')
library(ggrepel)
```

```
## Warning: package 'ggrepel' was built under R version 4.3.3
```

```
#install.packages('ggtext')
library(ggtext)
```

```
## Warning: package 'ggtext' was built under R version 4.3.3
```

```
# CO2 emissions data is mostly NULL from 2021 onwards...

indicators = c("SP.POP.TOTL", "EN.ATM.CO2E.PC", "NY.GDP.PCAP.KD")

wdi_co2_raw <- WDI(country = "all", indicators, extra = TRUE,
                   start = 1995, end = 2023)
```

```
wdi_clean <- wdi_co2_raw |>
  filter(region != "Aggregates") |>
  select(iso2c, iso3c, country, year,
    population = SP.POP.TOTL,
    co2_emissions = EN.ATM.CO2E.PC,
    gdp_per_cap = NY.GDP.PCAP.KD,
    region, income
  ) |>
  filter(population > 200000)
```

```
co2_rankings <- wdi_clean |>
  # Get rid of all the rows that have missing values in co2_emissions
  drop_na(co2_emissions) |>
  # Look at each year individually and rank countries based on their emissions that year
  mutate(
    ranking = rank(co2_emissions),
    .by = year
  )
```

Task 1: Prepare data in wide format

```
# YOUR CODE HERE

# seems like we only need the ranking columns
co2_rankings <- subset(co2_rankings, select = c("iso3c", "country", "region", "income", "ranking", "yea

co2_rankings <- reshape(co2_rankings, v.names = c("ranking"), idvar = "iso3c", timevar = "year", direct
```

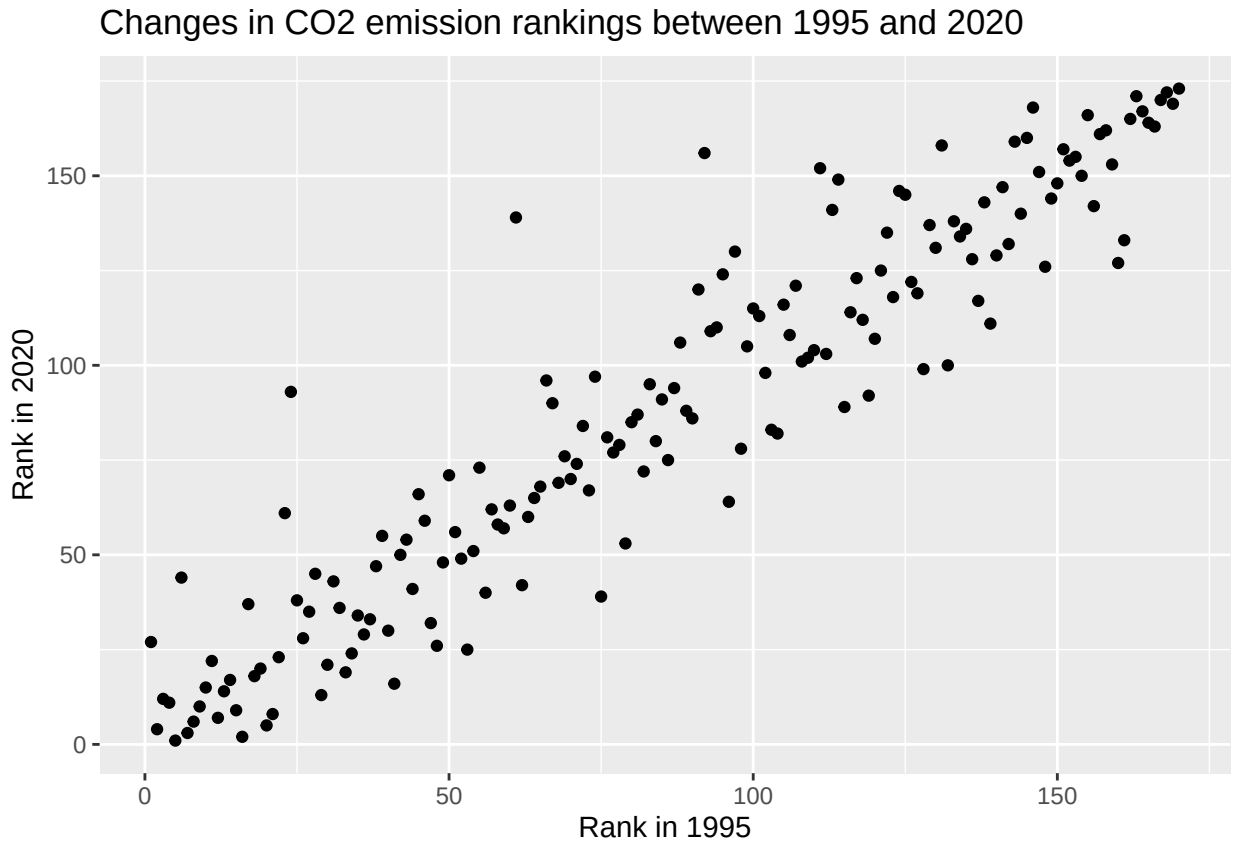
Task 2: Data wrangling

```
# YOUR CODE HERE

# Assuming the difference is between 1995 and 2020, based on the later tasks.
co2_rankings$rank_diff <- co2_rankings$ranking_2020 - co2_rankings$ranking_1995

co2_rankings <- co2_rankings %>% mutate(significant_diff = case_when(
  rank_diff > 30 ~ "Significant decrease",
  rank_diff < -30 ~ "Significant increase",
  .default = "Insignificant change"
))
```

Task 3: Scatter plot for changes in CO2 emission rankings between 1995 and 2020

```
# YOUR CODE HERE

ggplot(co2_rankings, aes(x = ranking_1995, y = ranking_2020)) +
  geom_point() +
  labs(
    x = "Rank in 1995",
    y = "Rank in 2020",
    title = "Changes in CO2 emission rankings between 1995 and 2020"
  )
```

```
## Warning: Removed 3 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

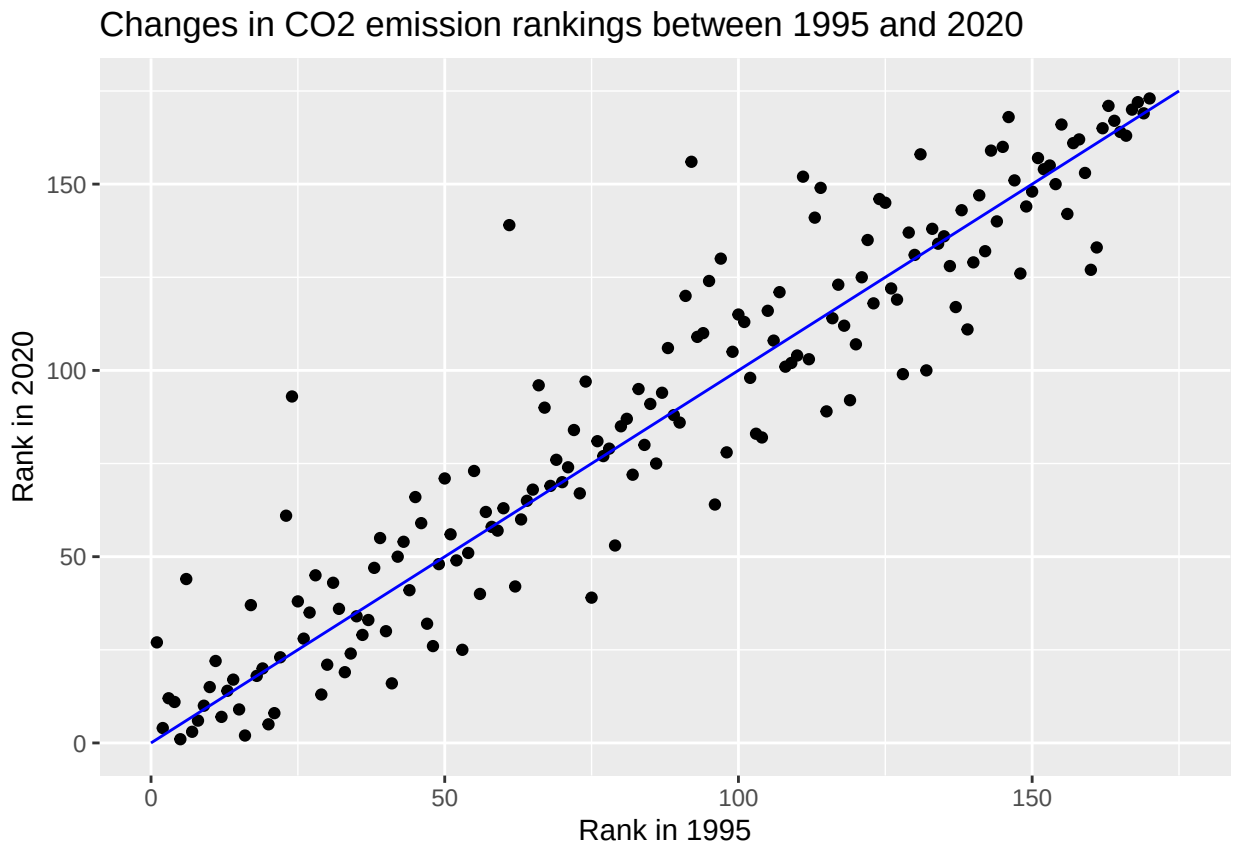


Task 4: Lazy way to show change in rank

```
# YOUR CODE HERE
```

```r
#points above the line => decreased rank, points below the line => increased rank, points on the line =
ggplot(co2_rankings, aes(x = ranking_1995, y = ranking_2020)) +
  geom_point() +
  labs(
    x = "Rank in 1995",
    y = "Rank in 2020",
    title = "Changes in CO2 emission rankings between 1995 and 2020"
  ) +
  annotate("segment", x = 0, xend = 175, y = 0, yend = 175, colour = "blue")
```

```
## Warning: Removed 3 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



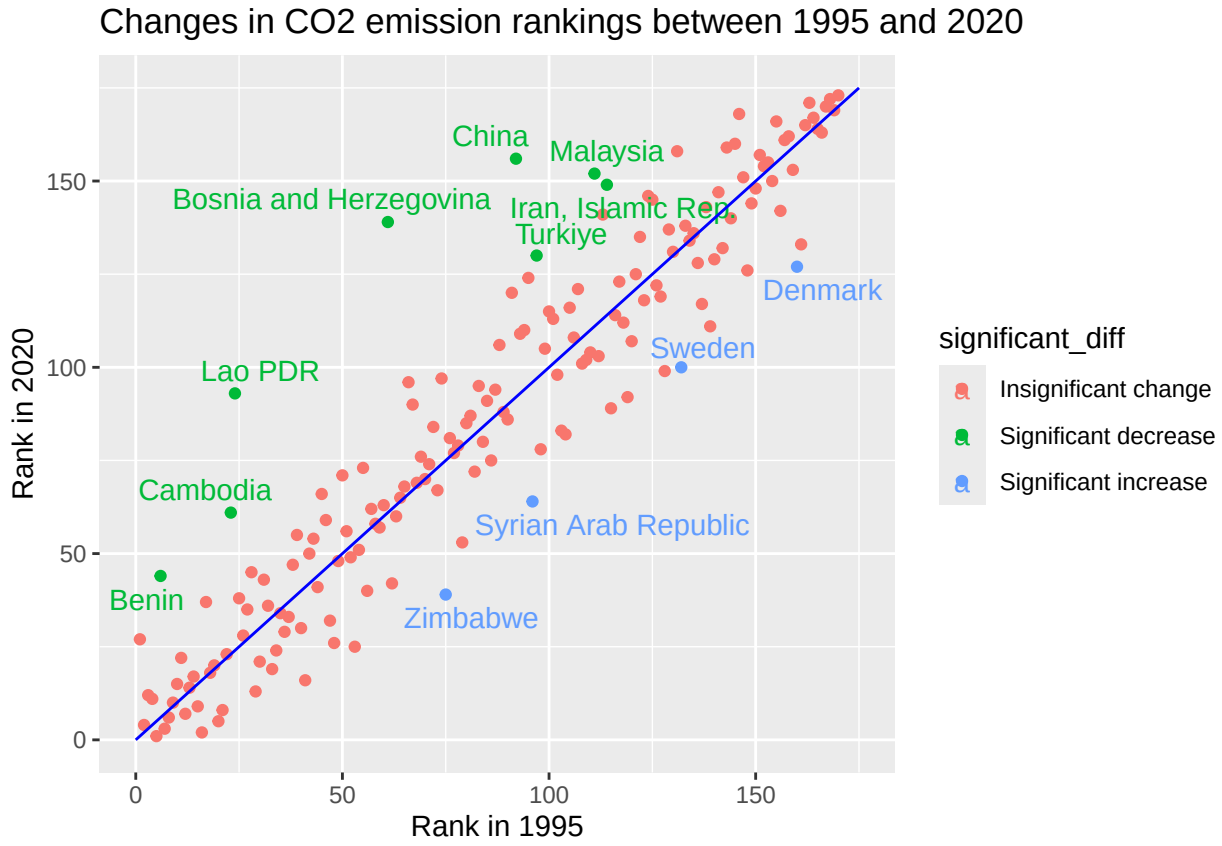
Task 5: Highlight significant countries

```r
# YOUR CODE HERE

co2_rankings <- co2_rankings %>% mutate(name_label = case_when(
  rank_diff > 30 | rank_diff < - 30 ~ co2_rankings$country,
  .default = ""
))
```

```
ggplot(co2_rankings, aes(x = ranking_1995, y = ranking_2020, color = significant_diff, label = name_lab
  geom_point() +
  geom_text_repel() +
  labs(
    x = "Rank in 1995",
    y = "Rank in 2020",
    title = "Changes in CO2 emission rankings between 1995 and 2020"
  ) +
  annotate("segment", x = 0, xend = 175, y = 0, yend = 175, colour = "blue")
```

```
## Warning: Removed 3 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

```
## Warning: Removed 3 rows containing missing values or values outside the scale range
## (`geom_text_repel()`).
```



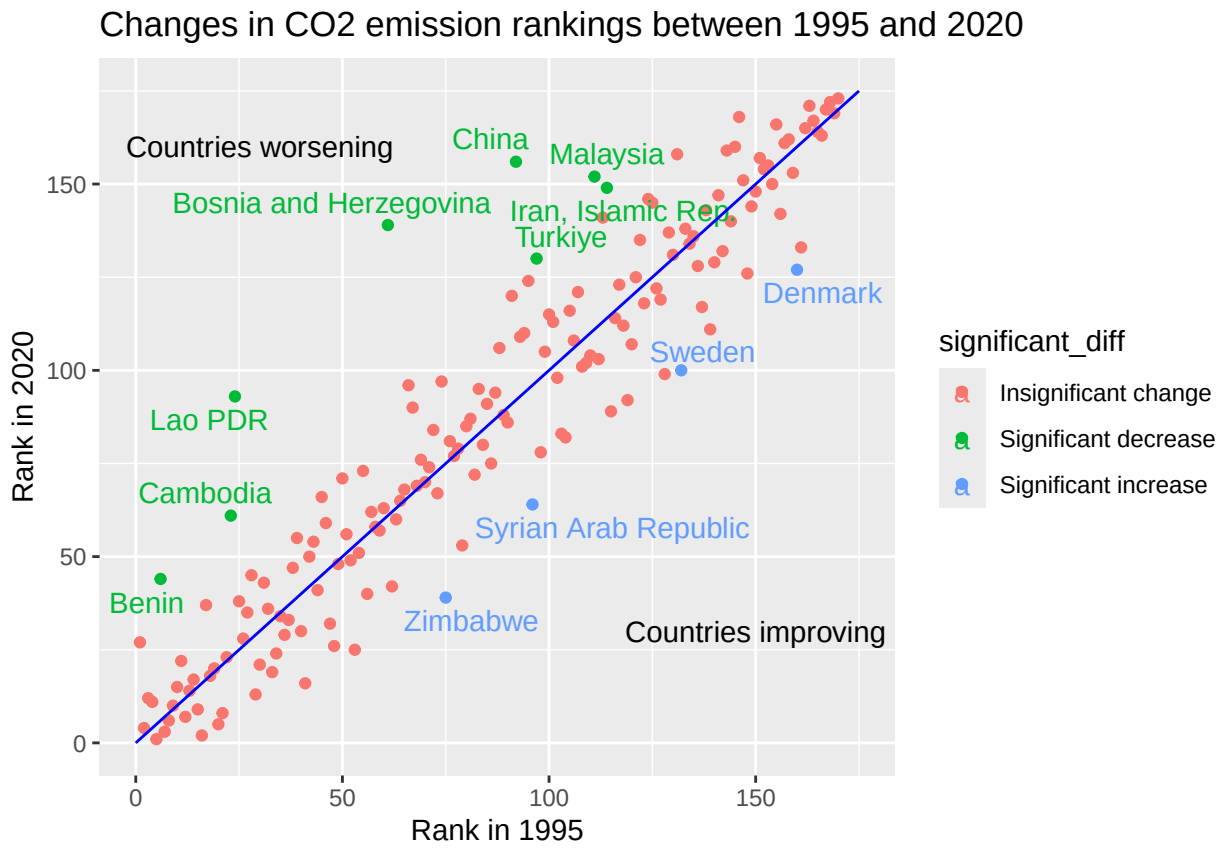
Task 6: Additional text annotations

```
# YOUR CODE HERE

co2_rankings <- co2_rankings %>% mutate(name_label = case_when(
```

```
  rank_diff > 30 | rank_diff < - 30 ~ co2_rankings$country,
  .default = ""
))

ggplot(co2_rankings, aes(x = ranking_1995, y = ranking_2020, color = significant_diff, label = name_lab
  geom_point() +
  geom_text_repel() +
  labs(
    x = "Rank in 1995",
    y = "Rank in 2020",
    title = "Changes in CO2 emission rankings between 1995 and 2020"
  ) +
  annotate("segment", x = 0, xend = 175, y = 0, yend = 175, colour = "blue") +
  annotate("text", x = 30, y = 160, label = "Countries worsening") +
  annotate("text", x = 150, y = 30, label = "Countries improving")
```

```
## Warning: Removed 3 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

```
## Warning: Removed 3 rows containing missing values or values outside the scale range
## (`geom_text_repel()`).
```

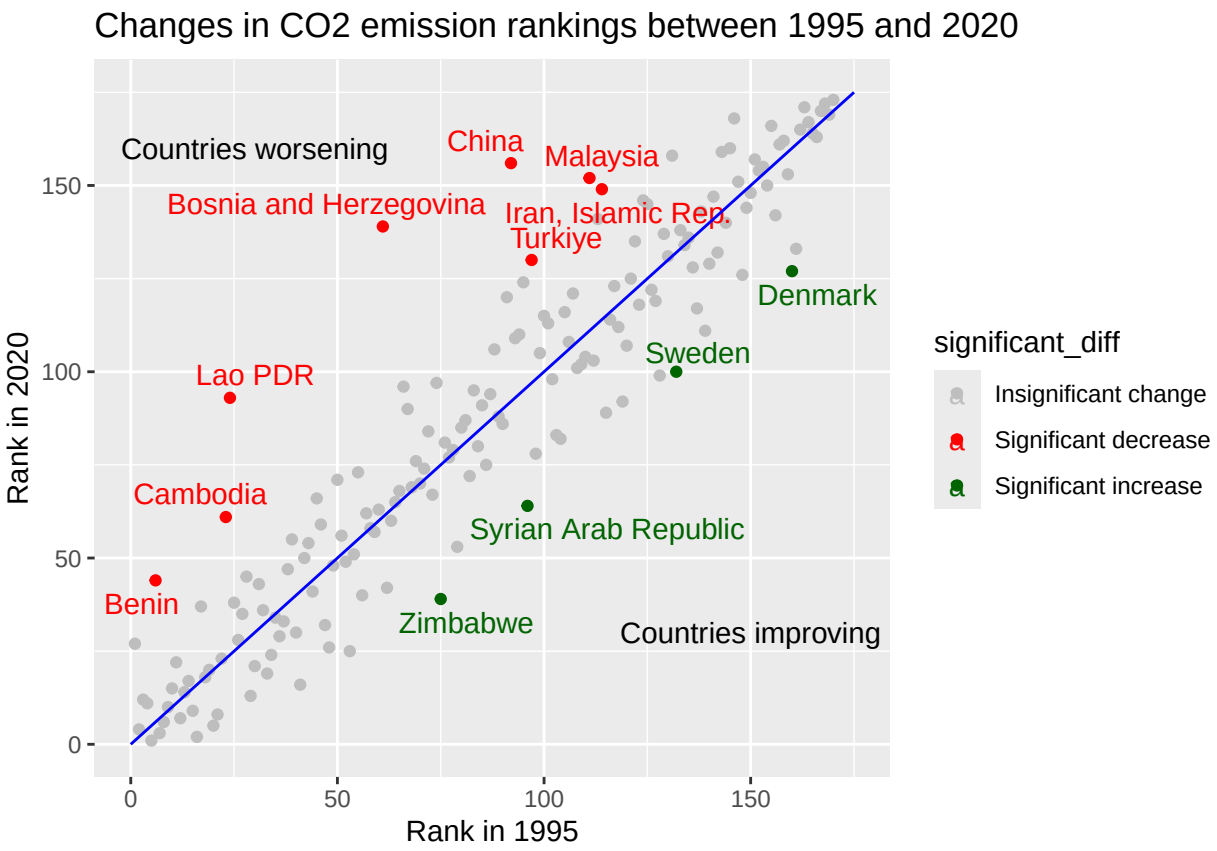


Task 7: Using colors to redirect attention

```
# YOUR CODE HERE

co2_rankings <- co2_rankings %>% mutate(name_label = case_when(
  rank_diff > 30 | rank_diff < - 30 ~ co2_rankings$country,
  .default = ""
))

ggplot(co2_rankings, aes(x = ranking_1995, y = ranking_2020, color = significant_diff, label = name_labe
  geom_point() +
  geom_text_repel() +
  labs(
    x = "Rank in 1995",
    y = "Rank in 2020",
    title = "Changes in CO2 emission rankings between 1995 and 2020"
  ) +
  annotate("segment", x = 0, xend = 175, y = 0, yend = 175, colour = "blue") +
  annotate("text", x = 30, y = 160, label = "Countries worsening") +
  annotate("text", x = 150, y = 30, label = "Countries improving") +
  scale_color_manual(values = c("gray", "red", "darkgreen"))
```

```
## Warning: Removed 3 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

```
## Warning: Removed 3 rows containing missing values or values outside the scale range
## (`geom_text_repel()`).
```



Task 8: More geometric annotations

```
# YOUR CODE HERE

co2_rankings <- co2_rankings %>% mutate(name_label = case_when(
  rank_diff > 30 | rank_diff < - 30 ~ co2_rankings$country,
  .default = ""
))

ggplot(co2_rankings, aes(x = ranking_1995, y = ranking_2020, color = significant_diff, label = name_lab
  geom_point() +
  geom_text_repel() +
  labs(
    x = "Rank in 1995",
    y = "Rank in 2020",
    title = "Changes in CO2 emission rankings between 1995 and 2020"
  ) +
  annotate("segment", x = 0, xend = 175, y = 0, yend = 175, colour = "blue") +
  annotate("text", x = 30, y = 160, label = "Countries worsening") +
  annotate("text", x = 150, y = 30, label = "Countries improving") +
  annotate("rect", xmin = 0, xmax = 25, ymin = 0, ymax = 25, alpha = .2) +
  annotate("segment", x = 50, xend = 25, y = 13, yend = 13, size = 1.5, arrow = arrow()) +
  annotate("text", x = 71, y = 13, label = "Lowest emitters") +
```

```
  annotate("rect", xmin = 148, xmax = 173, ymin = 148, ymax = 173, alpha = .2) +
  annotate("segment", x = 125, xend = 150, y = 160, yend = 160, size = 1.5, arrow = arrow()) +
  annotate("text", x = 123, y = 174, label = "Highest emitters") +
  scale_color_manual(values = c("gray", "red", "darkgreen"))
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

```
## Warning: Removed 3 rows containing missing values or values outside the scale range
## (‘geom_point()‘).
```

```
## Warning: Removed 3 rows containing missing values or values outside the scale range
## (‘geom_text_repel()‘).
```