# COMP4010 - Week 4

2024-03-14

## Week 4

## Application Exercises

```r
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
```

```r
library(ggwaffle)
library(emojifont)
```

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

```r
df <- read.csv("worldometer_data.csv", na.strings = "") #Other people didn't seem to need the na.string
```

### Task 1.

```r
ggplot(
  data = df,
  mapping = aes(x = TotalCases, y = TotalDeaths, color = Continent)
) +
geom_point() +
labs(
  x = "Total Cases", y = "Total Deaths",
  title = "Total Cases vs Total Deaths"
) +
theme(plot.title = element_text(hjust = 0.5))
```

```
## Warning: Removed 21 rows containing missing values or values outside the scale range
## ('geom_point()').
```
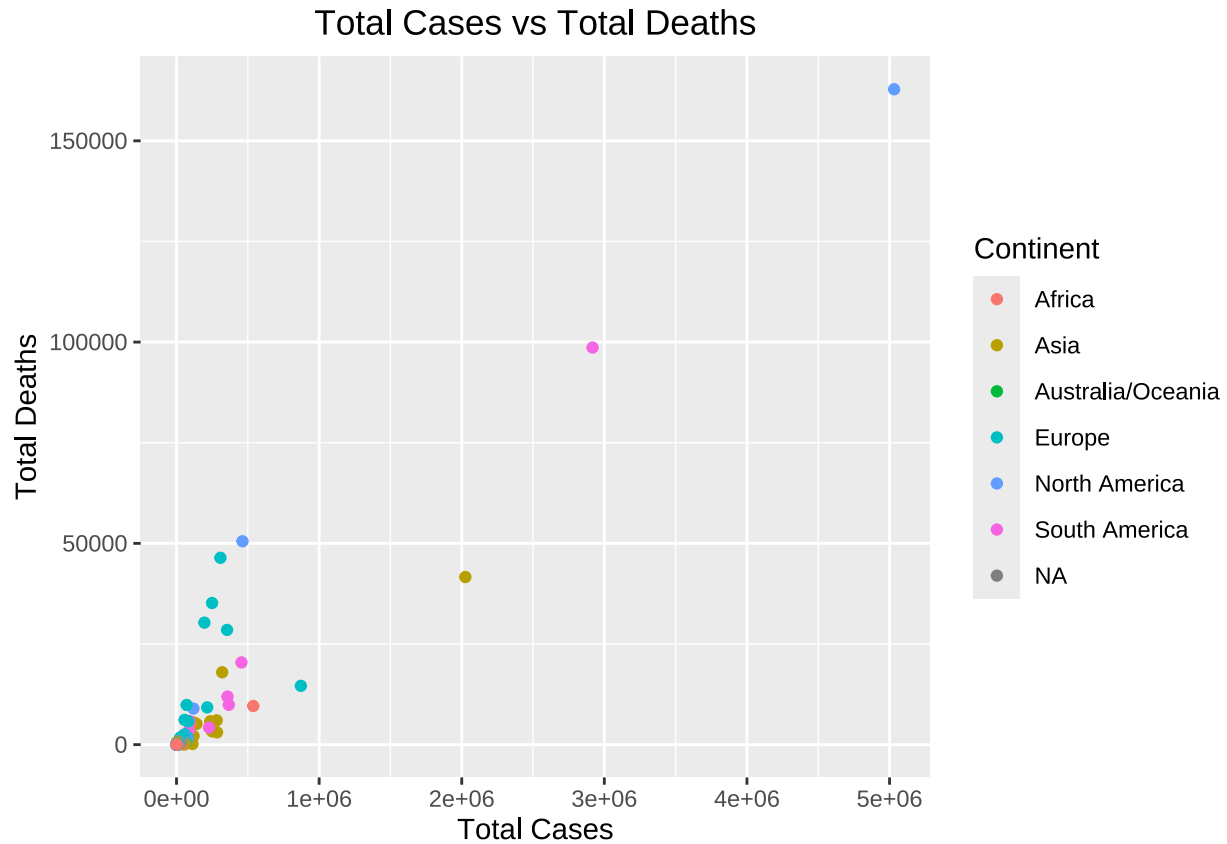


There's a positive correlation between the two variables - more total cases usually leads to higher death tolls. The extreme outliers seem to portray a linear relationship between the variables, while the remaining points seem to point towards something closer to an exponential function, although this is not clear due to cluttering and the two axes using different scales.

## Task 2.

```
ggplot(
  data = df,
  mapping = aes(x = TotalCases, y = TotalDeaths, color = Continent, scale(l))
  ) +
  geom_point() +
  scale_x_continuous(trans='log10') +
  scale_y_continuous(trans='log10') +
  labs(
    x = "Total Cases", y = "Total Deaths",
    title = "Total Cases vs Total Deaths"
  ) +
  theme(plot.title = element_text(hjust = 0.5))
```

```
## Warning: Removed 21 rows containing missing values or values outside the scale range
```

```
## (`geom_point()`).
```

## Total Cases vs Total Deaths
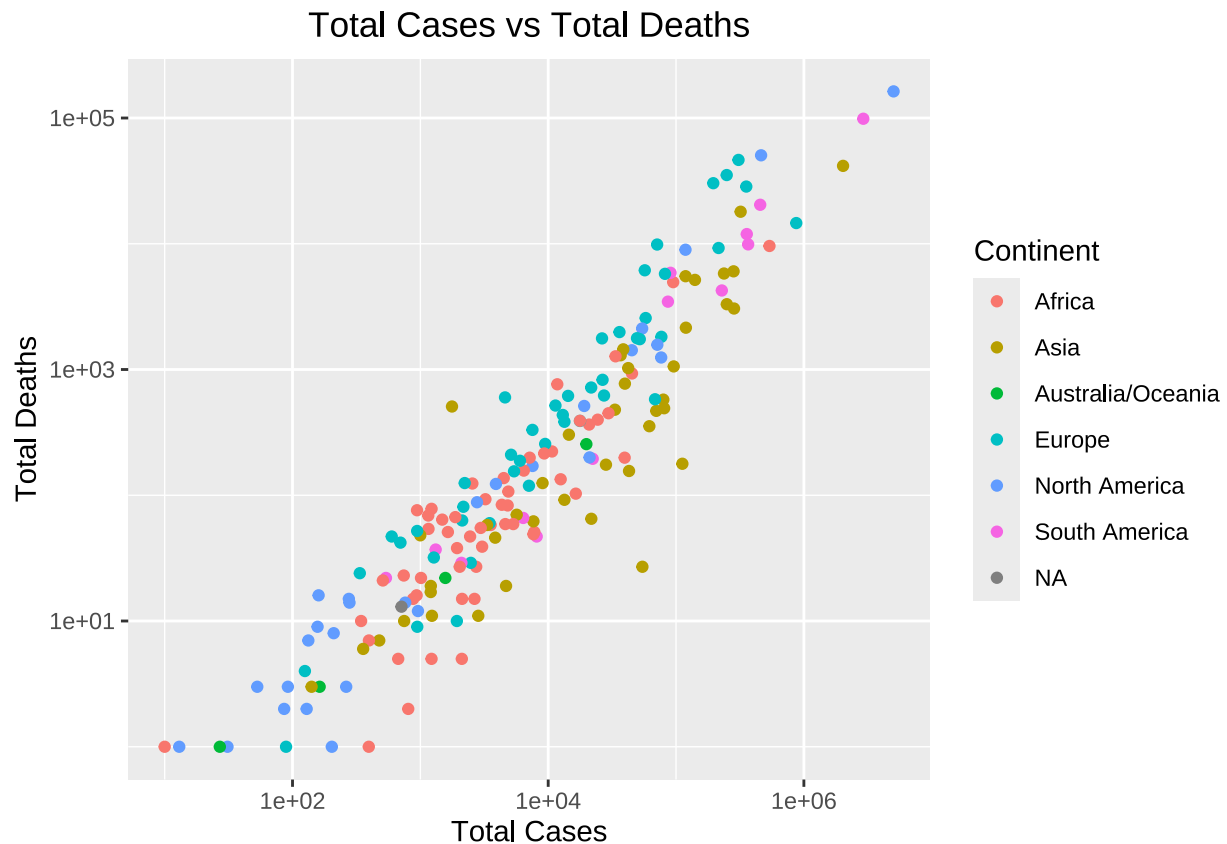


Before transforming the scale, a small number of outliers led to all other points being cluttered, making it hard to view the relationship between the two variables; the difference between the scales used for each axis also seems to exaggerate the relationship between total cases and total deaths. Using log scale brings the outliers closer the remaining data points, allowing the plot to clearly visualize the linear relationship between the variables.
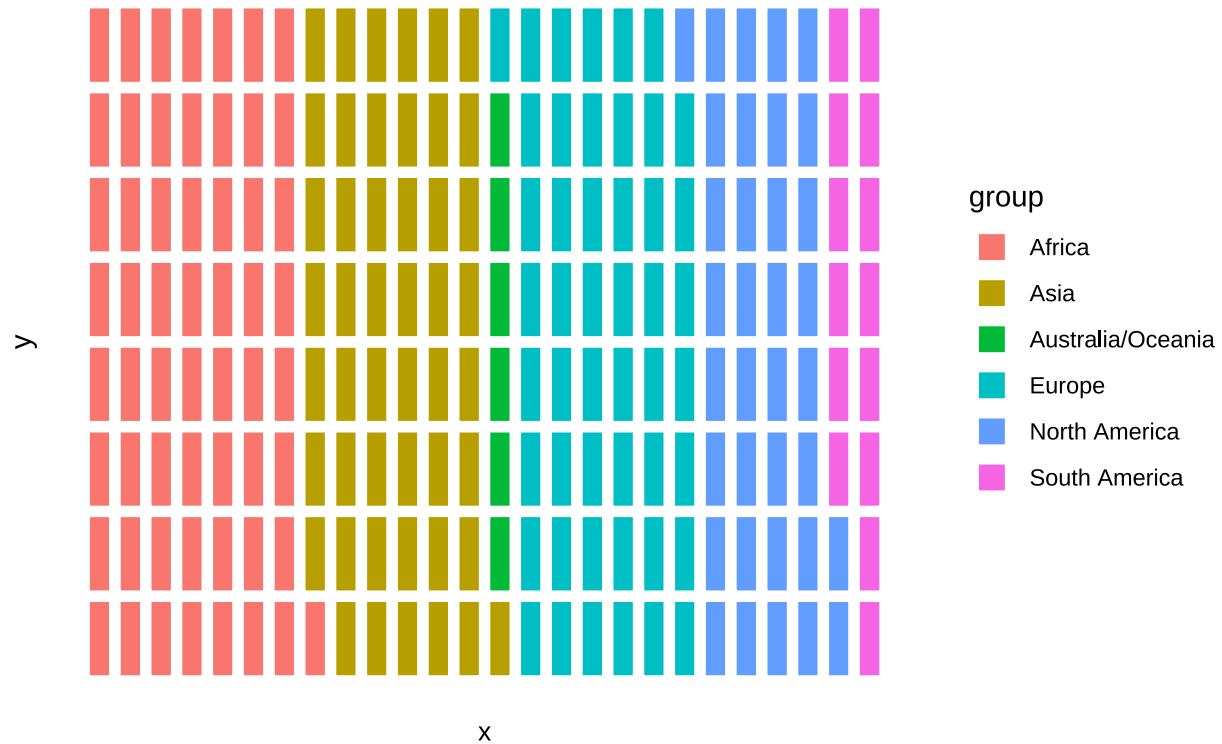
## Task 3.

```
waffle_data <- waffle_iron(df, aes_d(group = Continent))

ggplot(waffle_data, aes(x, y, fill = group)) +
  geom_waffle() +
  theme_waffle() +
  labs(
    title = "Worldometer by Continent"
  ) +
  theme(plot.title = element_text(hjust = 0.5))
```

```
## Warning: Using the `size` aesthetic in this geom was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` in the `default_aes` field and elsewhere instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
```

```
## generated.
```
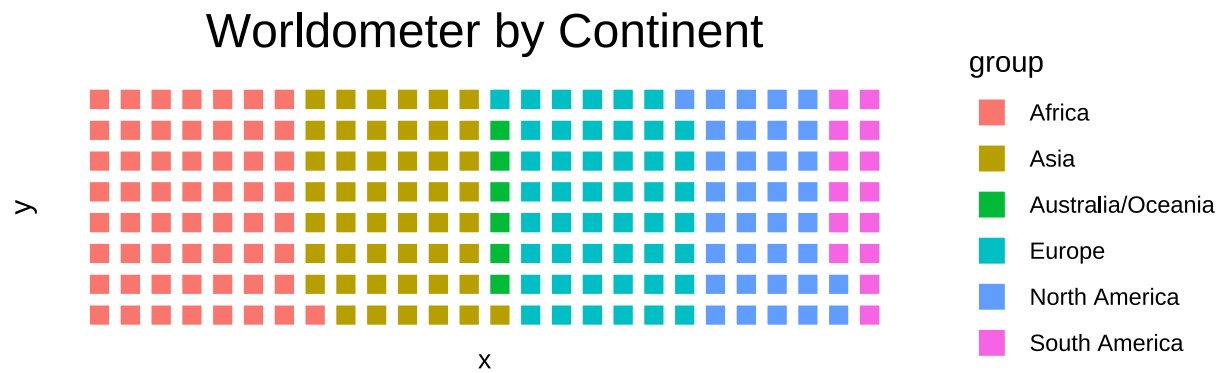
# Worldometer by Continent



Africa has the largest number of countries. Asia and Europe follow closely behind with 1 fewer country. Oceania has the fewest country by far with only 6.

## Task 4.

```r
waffle_data <- waffle_iron(df, aes_d(group = Continent))

ggplot(waffle_data, aes(x, y, fill = group)) +
  geom_waffle() +
  theme_waffle() +
  coord_equal() +
  labs(
    title = "Worldometer by Continent"
  ) +
  theme(plot.title = element_text(hjust = 0.5))
```
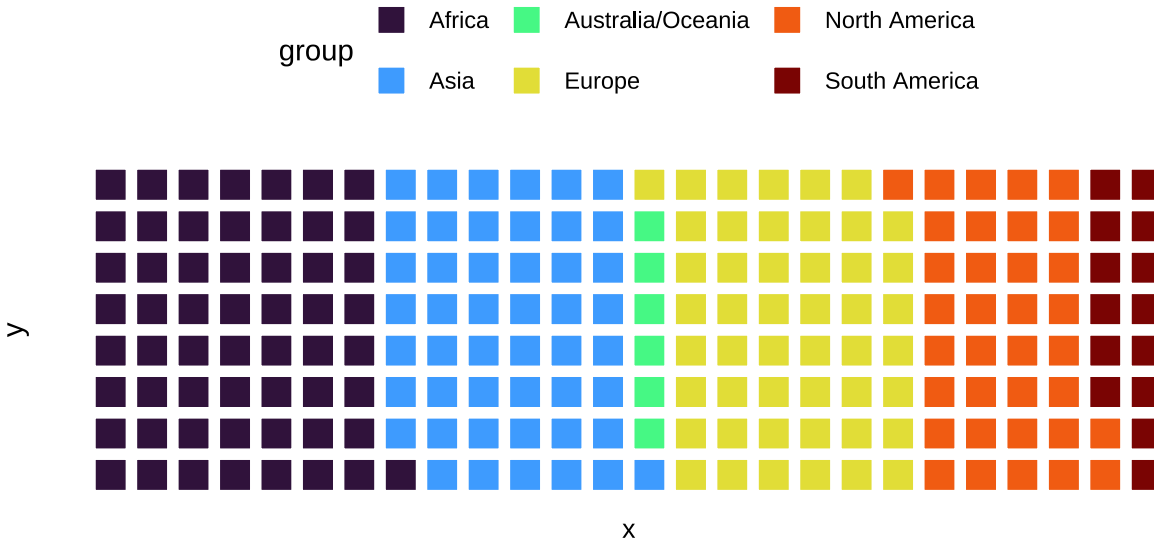
# Worldometer by Continent



**Task 5.**

```r
waffle_data <- waffle_iron(df, aes_d(group = Continent))

ggplot(waffle_data, aes(x, y, fill = group)) +
  geom_waffle() +
  theme_waffle() +
  scale_fill_viridis_d(option='turbo')+
  coord_equal() +
  labs(
    title = "Worldometer by Continent"
  ) +
  theme(plot.title = element_text(hjust = 0.5), legend.position = "top")
```

# Worldometer by Continent

group

- ■ Africa
- ■ Australia/Oceania
- ■ North America
- ■ Asia
- ■ Europe
- ■ South America



## Task 6.

```r
waffle_data <- waffle_iron(df, aes_d(group = Continent))

waffle_data$label = fontawesome('fa-globe')

ggplot(waffle_data, aes(x, y, colour = group)) +
  geom_text(aes(label=label), family='fontawesome-webfont', size=7) +
  theme_waffle() +
  coord_equal() +
  labs(
    title = "Worldometer by Continent"
  ) +
  theme(plot.title = element_text(hjust = 0.5), legend.position = "top")
```

# Worldometer by Continent

group

Africa
Australia/Oceania
North America
Asia
Europe
South America

y

x