World Health Organisation: Global Suicide Trends & Analysis

Created by: SKY_TKP

08 Sep 2024

Contents

Introduction

I'm 21 and currently study as Chulalongkorn University in Thailand.

This is my first large project with R programming. I'm proud it!

If you think something can be improving, then you can advise me via my gitbub

Let's enjoy ^_^

Data

This compiled dataset pulled from four other datasets linked by time and place, and was built to find signals correlated to increased suicide rates among different cohorts globally, across the socio-economic spectrum.

References

United Nations Development Program. (2018). Human development index (HDI). Retrieved from http://hdr.undp.org/en/indicators/137506

World Bank. (2018). World development indicators: GDP (current US$) by country:1985 to 2016. Retrieved from http://databank.worldbank.org/data/source/world-development-indicators#

[Szamil]. (2017). Suicide in the Twenty-First Century [dataset]. Retrieved from https://www.kaggle.com/szamil/suicide-in-the-twenty-first-century/notebook

World Health Organization. (2018). Suicide prevention. Retrieved from http://www.who.int/mental_health/suicide-prevention/en/

Import & Tidy

The majority of the data used in this analysis was obtained from the World Health Organisation.

Detial for dataset: master.csv

1. YEAR: CHRISTIAN ERA [CONTINUOUS]
2. AGE: AGE HAS BEEN SUMMARIZED WITHIN SPECIFIC RANGE [CATEGORICAL]
3. SUICIDES_NO: NUMBER OF SUICIDES [CONTINUOUS]
4. SEX: MALE/FEMALE [CATEGORICAL]
5. GDP_FOR_YEAR: PER YEAR GDP IN DOLLARS [CONTINUOUS]
6. GDP_PER_CAPITA:PER CAPITA GDP IN DOLLARS [CONTINUOUS]
7. SUICIDES/100K POP: SUICIDE BY POPULATION OF 100000 PEOPLE [CONTNUOUS]
8. COUNTRY_YEAR:COMBINATION OF COUNTRY-YEAR [CATEGORICAL]
9. HDI FOR YEAR: THE HUMAN DEVELOPMENT INDEX IS A SIMPLE FACTOR FOR MEASURING THE AVERAGE DEGREE OF ACCOMPLISHMENT IN A COUNTRY REGARDING THREE ASPECTS OF HUMAN DEVELOPMENT: HEALTH, EDUCATION AND INCOME [CONTINUOUS]
10. GENERATION: GENERATIONS IN AMERICA [CATEGORICAL]
11. COUNTRY: 93 COUNTRIES IN DATASET [CATEGORICAL]
12. POPULATION: SHOWS POPULATION OF GIVEN COUNTRY [CONTINUOUS]

Data Cleaning Notes

- 7 countries removed (<= 3 years of data total)
- 2016 data was removed (few countries had any, those that did often had data missing)
- HDI was removed due to 2/3 missing data
- Generation variable has problems, detailed in 2.11
- Continent was added to the dataset using the `countrycode` package
- Africa has very few countries providing suicide data

Import Library ,and Dataset

```
library(tidyverse) # general
library(ggalt) # dumbbell plots
library(countrycode) # continent
library(rworldmap) # quick country-level heat maps
library(gridExtra) # plots
library(broom) # significant trends within countries

theme_set(theme_light())
```

About them (via MS Excel)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	country	year	sex	age	suicides_n	population	suicides/1(	country-ye	HDI for ye	gdp_for_year ($)	gdp_per_c	generation	
2	Albania	1987	male	15-24 year	21	312900	6.71	Albania1987		2,156,624,900	796	Generation X	
3	Albania	1987	male	35-54 year	16	308000	5.19	Albania1987		2,156,624,900	796	Silent	
4	Albania	1987	female	15-24 year	14	289700	4.83	Albania1987		2,156,624,900	796	Generation X	
5	Albania	1987	male	75+ years	1	21800	4.59	Albania1987		2,156,624,900	796	G.I. Generation	
6	Albania	1987	male	25-34 year	9	274300	3.28	Albania1987		2,156,624,900	796	Boomers	
7	Albania	1987	female	75+ years	1	35600	2.81	Albania1987		2,156,624,900	796	G.I. Generation	
8	Albania	1987	female	35-54 year	6	278800	2.15	Albania1987		2,156,624,900	796	Silent	
9	Albania	1987	female	25-34 year	4	257200	1.56	Albania1987		2,156,624,900	796	Boomers	
10	Albania	1987	male	55-74 year	1	137500	0.73	Albania1987		2,156,624,900	796	G.I. Generation	
11	Albania	1987	female	5-14 years	0	311000	0	Albania1987		2,156,624,900	796	Generation X	
12	Albania	1987	female	55-74 year	0	144600	0	Albania1987		2,156,624,900	796	G.I. Generation	
13	Albania	1987	male	5-14 years	0	338200	0	Albania1987		2,156,624,900	796	Generation X	
14	Albania	1988	female	75+ years	2	36400	5.49	Albania1988		2,126,000,000	769	G.I. Generation	
15	Albania	1988	male	15-24 year	17	319200	5.33	Albania1988		2,126,000,000	769	Generation X	
16	Albania	1988	male	75+ years	1	22300	4.48	Albania1988		2,126,000,000	769	G.I. Generation	
17	Albania	1988	male	35-54 year	14	314100	4.46	Albania1988		2,126,000,000	769	Silent	
18	Albania	1988	male	55-74 year	4	140200	2.85	Albania1988		2,126,000,000	769	G.I. Generation	
19	Albania	1988	female	15-24 year	8	295600	2.71	Albania1988		2,126,000,000	769	Generation X	
20	Albania	1988	female	55-74 year	3	147500	2.03	Albania1988		2,126,000,000	769	G.I. Generation	
21	Albania	1988	female	25-34 year	5	262400	1.91	Albania1988		2,126,000,000	769	Boomers	
22	Albania	1988	male	25-34 year	5	279900	1.79	Albania1988		2,126,000,000	769	Boomers	

```
# 1) Import & data cleaning

data <- read_csv("master.csv")
```

```
## Rows: 27820 Columns: 12
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): country, sex, age, country-year, generation
## dbl (6): year, suicides_no, population, suicides/100k pop, HDI for year, gdp...
## num (1): gdp_for_year ($)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
dim(data) # 27820 rows x 12 columns
```

```
## [1] 27820    12
```

```r
summary(data)
```

```
##    country               year          sex                age
##  Length:27820       Min.   :1985   Length:27820       Length:27820
##  Class :character   1st Qu.:1995   Class :character   Class :character
##  Mode  :character   Median :2002   Mode  :character   Mode  :character
##                     Mean   :2001
##                     3rd Qu.:2008
##                     Max.   :2016
##
##   suicides_no        population       suicides/100k pop country-year
##  Min.   :    0.0   Min.   :     278   Min.   :  0.00    Length:27820
##  1st Qu.:    3.0   1st Qu.:   97498   1st Qu.:  0.92    Class :character
##  Median :   25.0   Median :  430150   Median :  5.99    Mode  :character
##  Mean   :  242.6   Mean   : 1844794   Mean   : 12.82
##  3rd Qu.:  131.0   3rd Qu.: 1486143   3rd Qu.: 16.62
##  Max.   :22338.0   Max.   :43805214   Max.   :224.97
##
##   HDI for year   gdp_for_year ($)    gdp_per_capita ($)  generation
##  Min.   :0.483   Min.   :4.692e+07   Min.   :   251     Length:27820
##  1st Qu.:0.713   1st Qu.:8.985e+09   1st Qu.:  3447     Class :character
##  Median :0.779   Median :4.811e+10   Median :  9372     Mode  :character
##  Mean   :0.777   Mean   :4.456e+11   Mean   : 16866
##  3rd Qu.:0.855   3rd Qu.:2.602e+11   3rd Qu.: 24874
##  Max.   :0.944   Max.   :1.812e+13   Max.   :126352
##  NA's   :19456
```

```r
# glimpse(data) # will tidy up these variable names

# sum(is.na(data$`HDI for year`)) # remove, > 2/3 missing, not usable

# table(data$age, data$generation) # don't like this variable

data <- data %>%
  select(-c(`HDI for year`, `suicides/100k pop`)) %>%
  rename(gdp_for_year   = `gdp_for_year ($)`,         # rename
         gdp_per_capita = `gdp_per_capita ($)`,
         country_year   = `country-year`) %>%
  as.data.frame()

#-----------------------------------------------------------------------
# 2) OTHER ISSUES

# a) this SHOULD give 12 rows for every county-year combination (6 age bands * 2 genders):

# data %>%
#   group_by(country_year) %>%
#   count() %>%
#   filter(n != 12) # note: there appears to be an issue with 2016 data
# not only are there few countries with data, but those that do have data are incomplete

data <- data %>%
  filter(year != 2016) %>% # therefore exclude 2016 data
```

```r
  select(-country_year)


# b) excluding countries with <= 3 years of data:

minimum_years <- data %>%
  group_by(country) %>%
  summarize(rows = n(),
            years = rows / 12) %>%
  #I want to find for each county what have much data ?
  arrange(years)

data <- data %>%
  filter( !( country %in% minimum_years[minimum_years$years <= 3, ] ) )

# dim(data) # 27492 rows x 10 columns

# no other major data issues found yet

#-----------------------------------------------------------------
# 3) TIDYING DATAFRAME
data$age <- gsub(" years", "", data$age) # delete "years" in columns "age"
data$sex <- ifelse(data$sex == "male", "Male", "Female")


# getting continent data:
data$continent <- countrycode(sourcevar = data[, "country"],
                              origin = "country.name",
                              destination = "continent")

# Nominal factors
data_nominal <- c('country', 'sex', 'continent')
data[data_nominal] <- lapply(data[data_nominal], function(x){factor(x)})


# Making age ordinal
data$age <- factor(data$age,
                   ordered = T,
                   levels = c("5-14",
                              "15-24",
                              "25-34",
                              "35-54",
                              "55-74",
                              "75+"))

# Making generation ordinal
data$generation <- factor(data$generation,
                   ordered = T,
                   levels = c("G.I. Generation",
                              "Silent",
                              "Boomers",
                              "Generation X",
                              "Millenials",
```

```
                              "Generation Z"))

data <- as_tibble(data)


# the global rate over the time period will be useful:

global_average <- (sum(as.numeric(data$suicides_no)) /
                     sum(as.numeric(data$population))) * 10^5

# view the finalized data
glimpse(data)
```

```
## Rows: 27,660
## Columns: 10
## $ country        <fct> Albania, Albania, Albania, Albania, Albania, Albania, A~
## $ year           <dbl> 1987, 1987, 1987, 1987, 1987, 1987, 1987, 1987, 1987, 1~
## $ sex            <fct> Male, Male, Female, Male, Male, Female, Female, Female,~
## $ age            <ord> 15-24, 35-54, 15-24, 75+, 25-34, 75+, 35-54, 25-34, 55-~
## $ suicides_no    <dbl> 21, 16, 14, 1, 9, 1, 6, 4, 1, 0, 0, 0, 2, 17, 1, 14, 4,~
## $ population     <dbl> 312900, 308000, 289700, 21800, 274300, 35600, 278800, 2~
## $ gdp_for_year   <dbl> 2156624900, 2156624900, 2156624900, 2156624900, 2156624~
## $ gdp_per_capita <dbl> 796, 796, 796, 796, 796, 796, 796, 796, 796, 796, 796, ~
## $ generation     <ord> Generation X, Silent, Generation X, G.I. Generation, Bo~
## $ continent      <fct> Europe, Europe, Europe, Europe, Europe, Europe, Europe,~
```

Key Insights

- Suicide rates are **decreasing globally. (2.1)**
- Of those **countries** that show clear linear trends over time, **2/3 are decreasing. (2.5.2)**
- On average, **suicide rate increases with age. (2.4)**
- This remains true when controlling for continent in the Americas, Asia & Europe, but not for Africa & Oceania. **(2.8)**
- There is a *weak* **positive relationship between a countries GDP (per capita) and suicide rate. (2.10)**
- The **highest suicide rate** ever recorded in a demographic (for 1 year) is **225 (per 100k population). (2.12)**
- There is an **overrepresentation of men** in suicide deaths at every level of analysis (globally, at a continent and country level). Globally, the male rate is ~3.5x higher. **(2.3) (2.6) (2.7)**

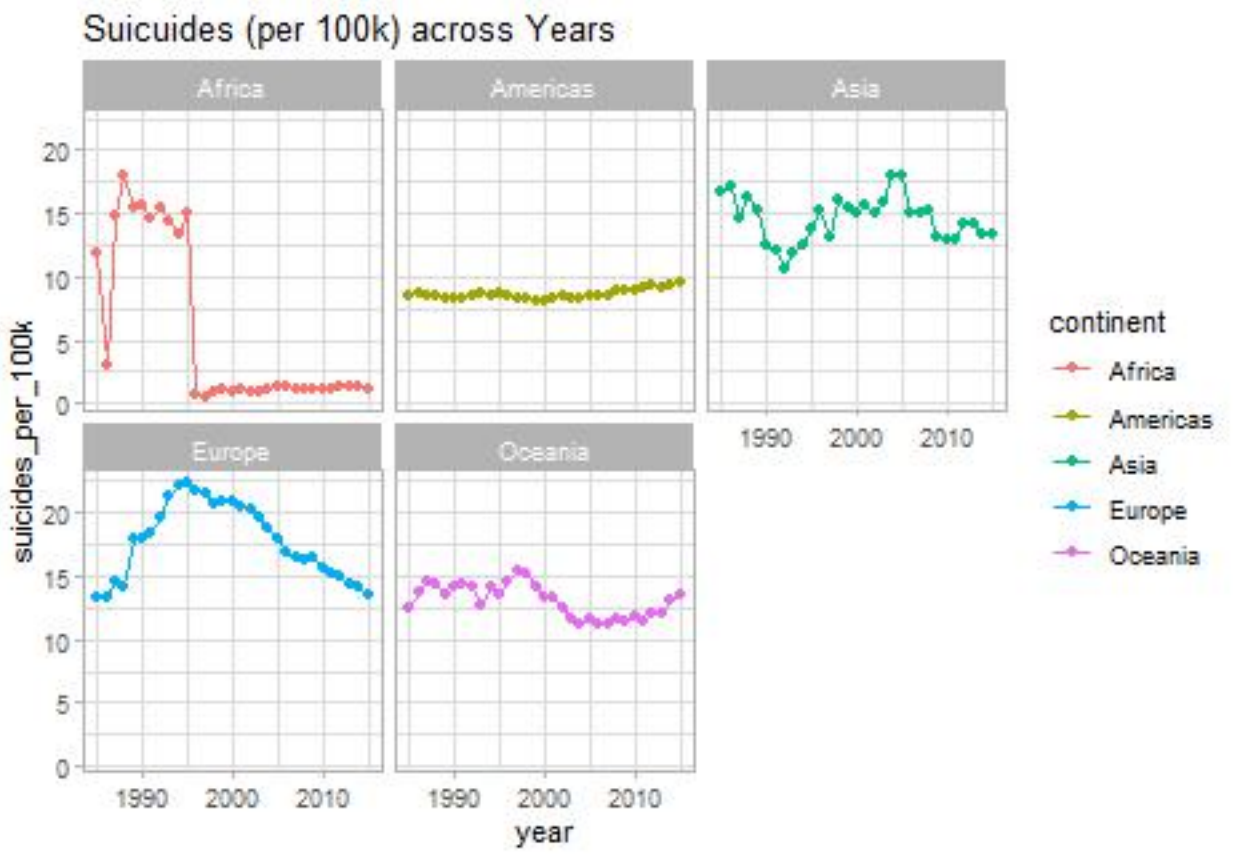
Global Analysis

Global Trend

The dashed line is the **global average suicide rate** from 1985 - 2015: **13.1 deaths** (per 100k, per year).

```
data %>%
  group_by(year, continent) %>%
  summarise(suicides_per_100k = sum(suicides_no) /
              sum(population) * 10^5 ) %>%
  ggplot(aes(x = year, y = suicides_per_100k, color = continent)) +
  geom_point(size = 1.5) +
  geom_line(alpha = 1.0) +
  facet_wrap(~ continent) +
  labs(title = "Suicuides (per 100k) across Years")
```

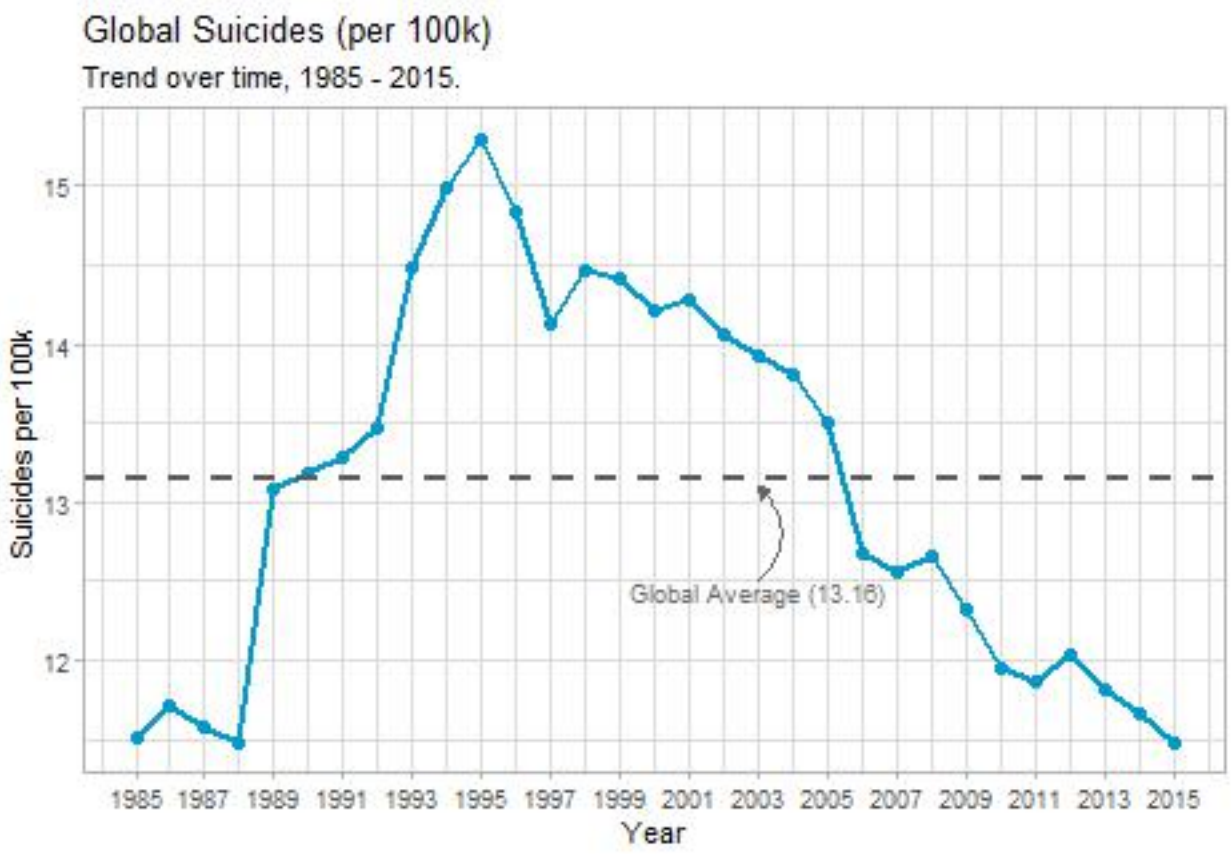
```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```




```
data %>%
  group_by(year) %>%
  summarize(population = sum(population),
            suicides = sum(suicides_no),
            suicides_per_100k = (suicides / population) * 10^5) %>%
```

```r
  ggplot(aes(x = year, y = suicides_per_100k)) +
  geom_line(col = "deepskyblue3", size = 1) +
  geom_point(col = "deepskyblue3", size = 2) +
  geom_hline(yintercept = global_average, linetype = 2, color = "grey35",
             size = 1) +
  annotate("text", x = 2003, y = 12.5,
           label = "Global Average (13.16)", vjust = 1,
           size = 3, color = "grey40") +
  annotate(
    "curve",
    x = 2003, y = 12.5,
    xend = 2003, yend = 13.1,
    arrow = arrow(length = unit(0.2, "cm"), type = "closed"),
    color = "grey40"
  )+
  labs(title = "Global Suicides (per 100k)",
       subtitle = "Trend over time, 1985 - 2015.",
       x = "Year",
       y = "Suicides per 100k") +
  scale_x_continuous(breaks = seq(1985, 2015, 2)) +
  scale_y_continuous(breaks = seq(10, 20))
```

```
## 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.
```



Insights

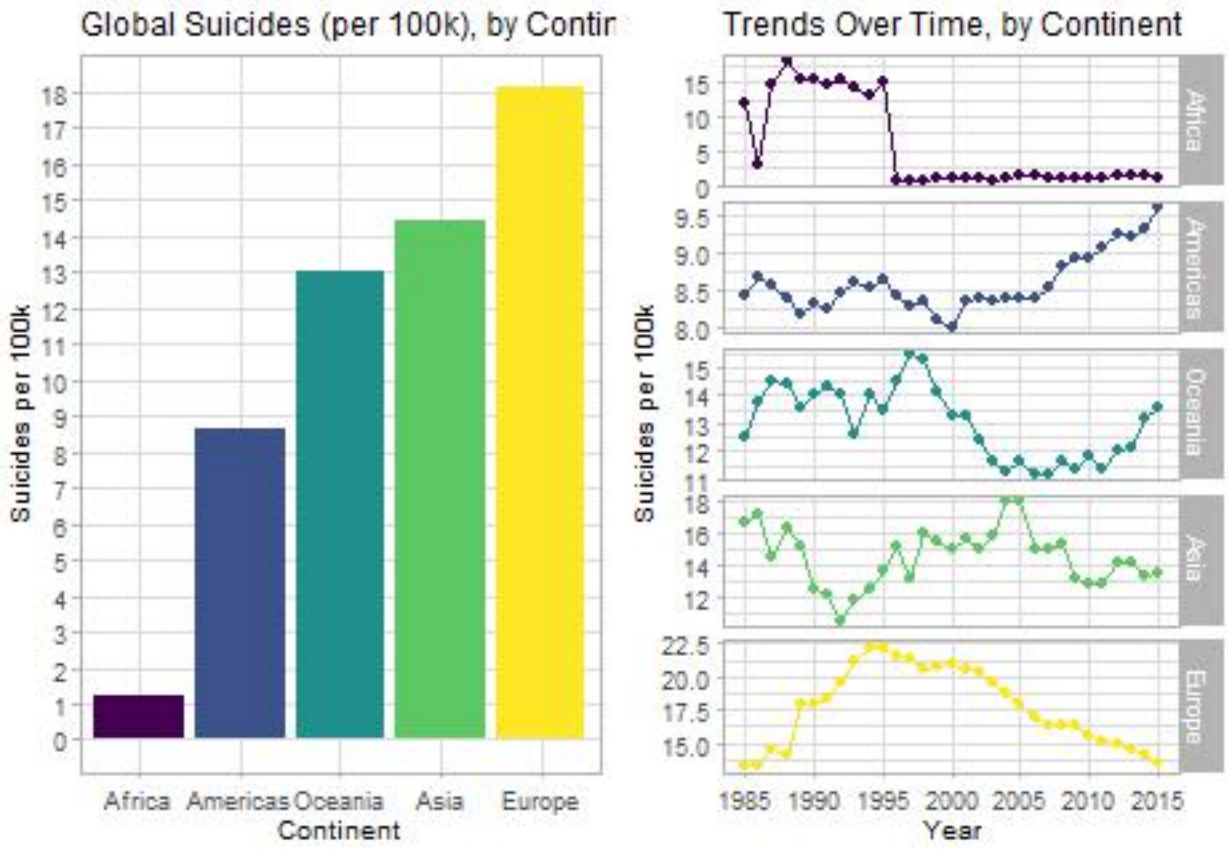
- Peak suicide rate was **15.3** deaths per 100k in **1995**
- Decreased steadily, to **11.5** per 100k in **2015** (~**25% decrease**)
- Rates are only now returning to their pre-90's rates
- **Limited data in the 1980's**, so it's hard to say if rate then was truly representative of the global population

By Continent

```
continent <- data %>%
  group_by(continent) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5) %>%
  arrange(suicide_per_100k)

continent$continent <- factor(continent$continent, ordered = T,
                              levels = continent$continent)

continent_plot <- ggplot(continent, aes(x = continent, y = suicide_per_100k, fill = continent)) +
  geom_bar(stat = "identity") +
  labs(
    title = "Global Suicides (per 100k), by Continent",
        x = "Continent",
        y = "Suicides per 100k",
    fill = "Continent") +
  theme(legend.position = "none",
        title = element_text(size = 10)) +
  scale_y_continuous(breaks = seq(0, 20, 1),
                     minor_breaks = F)


continent_time <- data %>%
  group_by(year, continent) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5)
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```

```
continent_time$continent <- factor(continent_time$continent, ordered = T
                                   , levels = continent$continent)

continent_time_plot <- ggplot(continent_time,
                              aes(x = year, y = suicide_per_100k,
                                  col = factor(continent))) +
  facet_grid(continent ~ ., scales = "free_y") +
  geom_line() +
  geom_point() +
  labs(title = "Trends Over Time, by Continent",
       x = "Year",
       y = "Suicides per 100k",
       color = "Continent") +
  theme(legend.position = "none", title = element_text(size = 10)) +
  scale_x_continuous(breaks = seq(1985, 2015, 5),
                     minor_breaks = F)

grid.arrange(continent_plot, continent_time_plot, ncol = 2)
```



Insights

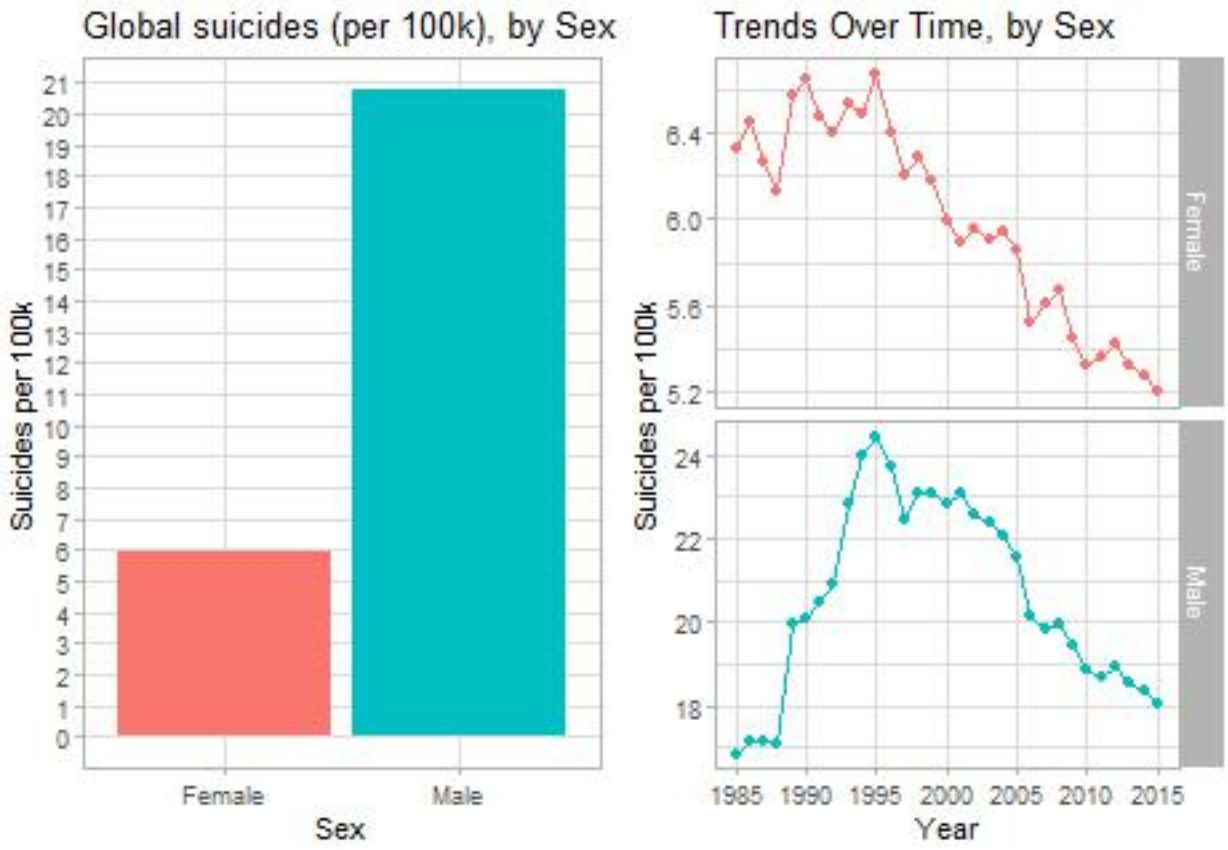
- **European rate highest overall**, but has steadily **decreasing since 1995**
- The **European rate for 2015 similar to Asia & Oceania**
- The trend-line for Africa is due to poor data quality - just 3 countries have provided data
- **Oceania & Americas trends are more concerning**

By Sex

```
sex_plot <- data %>%
  group_by(sex) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5) %>%
ggplot(aes(x = sex, y = suicide_per_100k, fill = sex)) +
  geom_bar(stat = "identity") +
  labs(title = "Global suicides (per 100k), by Sex",
       x = "Sex",
       y = "Suicides per 100k") +
  theme(legend.position = "none") +
  scale_y_continuous(breaks = seq(0, 25), minor_breaks = F)

### with time
sex_time_plot <- data %>%
  group_by(year, sex) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5) %>%
  ggplot(aes(x = year, y = suicide_per_100k, col = factor(sex))) +
  facet_grid(sex ~ ., scales = "free_y") +
  geom_line() +
  geom_point() +
  labs(title = "Trends Over Time, by Sex",
       x = "Year",
       y = "Suicides per 100k",
       color = "Sex") +
  theme(legend.position = "none") +
  scale_x_continuous(breaks = seq(1985, 2015, 5), minor_breaks = F)
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```

```
grid.arrange(sex_plot, sex_time_plot, ncol = 2)
```



Insights

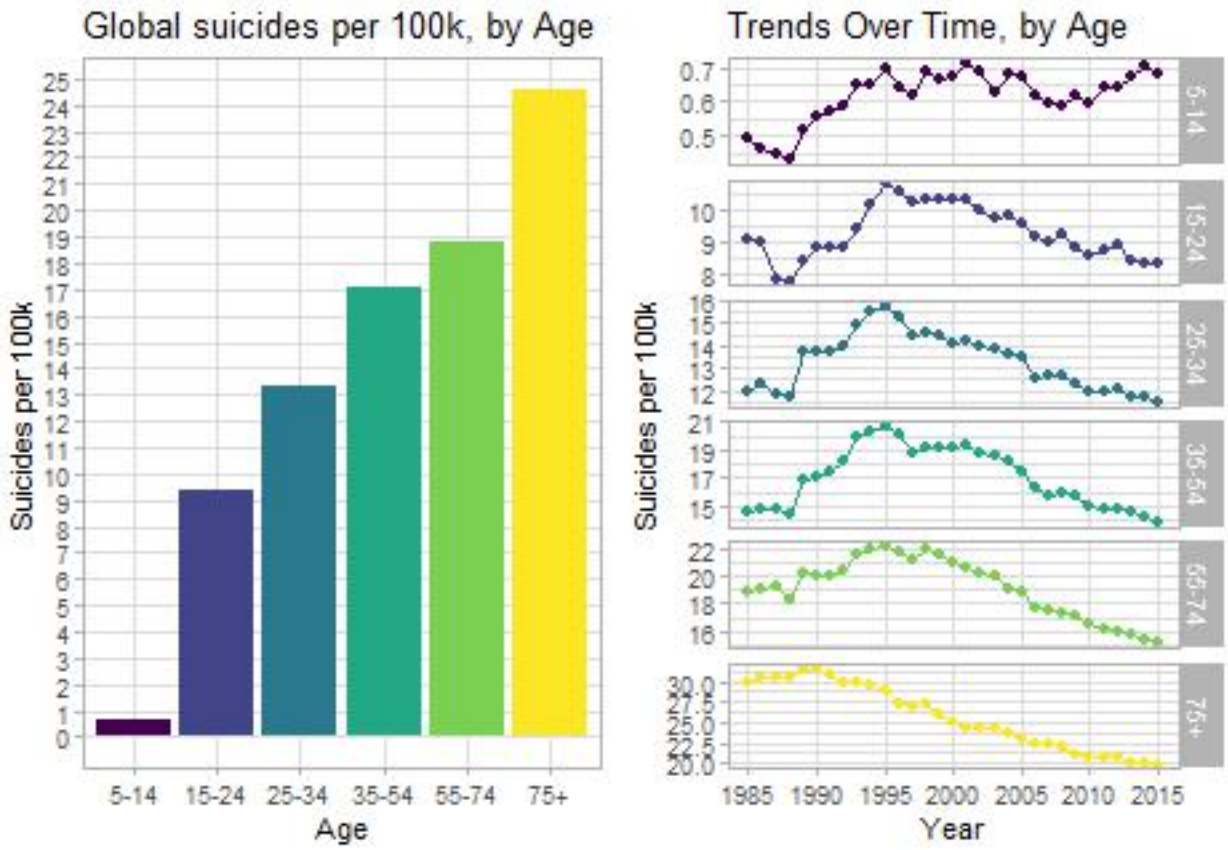
- Globally, the rate of suicide for men has been ~**3.5x higher for men**
- Both **male & female suicide rates peaked in 1995**, declining since
- This ratio of 3.5 : 1 (male : female) has remained relatively constant since the mid 90's
- However, during the 80's this ratio was as low as 2.7 : 1 (male : female)

By Age

```
age_plot <- data %>%
  group_by(age) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5) %>%
  ggplot(aes(x = age, y = suicide_per_100k, fill = age)) +
  geom_bar(stat = "identity") +
  labs(title = "Global suicides per 100k, by Age",
       x = "Age",
       y = "Suicides per 100k") +
  theme(legend.position = "none") +
  scale_y_continuous(breaks = seq(0, 30, 1), minor_breaks = F)

### with time
age_time_plot <- data %>%
  group_by(year, age) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5) %>%
  ggplot(aes(x = year, y = suicide_per_100k, col = age)) +
  facet_grid(age ~ ., scales = "free_y") +
  geom_line() +
  geom_point() +
  labs(title = "Trends Over Time, by Age",
       x = "Year",
       y = "Suicides per 100k",
       color = "Age") +
  theme(legend.position = "none") +
  scale_x_continuous(breaks = seq(1985, 2015, 5), minor_breaks = F)
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```

```
grid.arrange(age_plot, age_time_plot, ncol = 2)
```



Insights

- Globally, the **likelihood of suicide increases with age**
- Since 1995, suicide rate for everyone aged **>= 15** has been **linearly decreasing**
- The suicide rate of those aged **75+** has dropped by more than **50%** since 1990
- Suicide rate in the **'5-14'** category remains **roughly static** and small (< 1 per 100k per year)

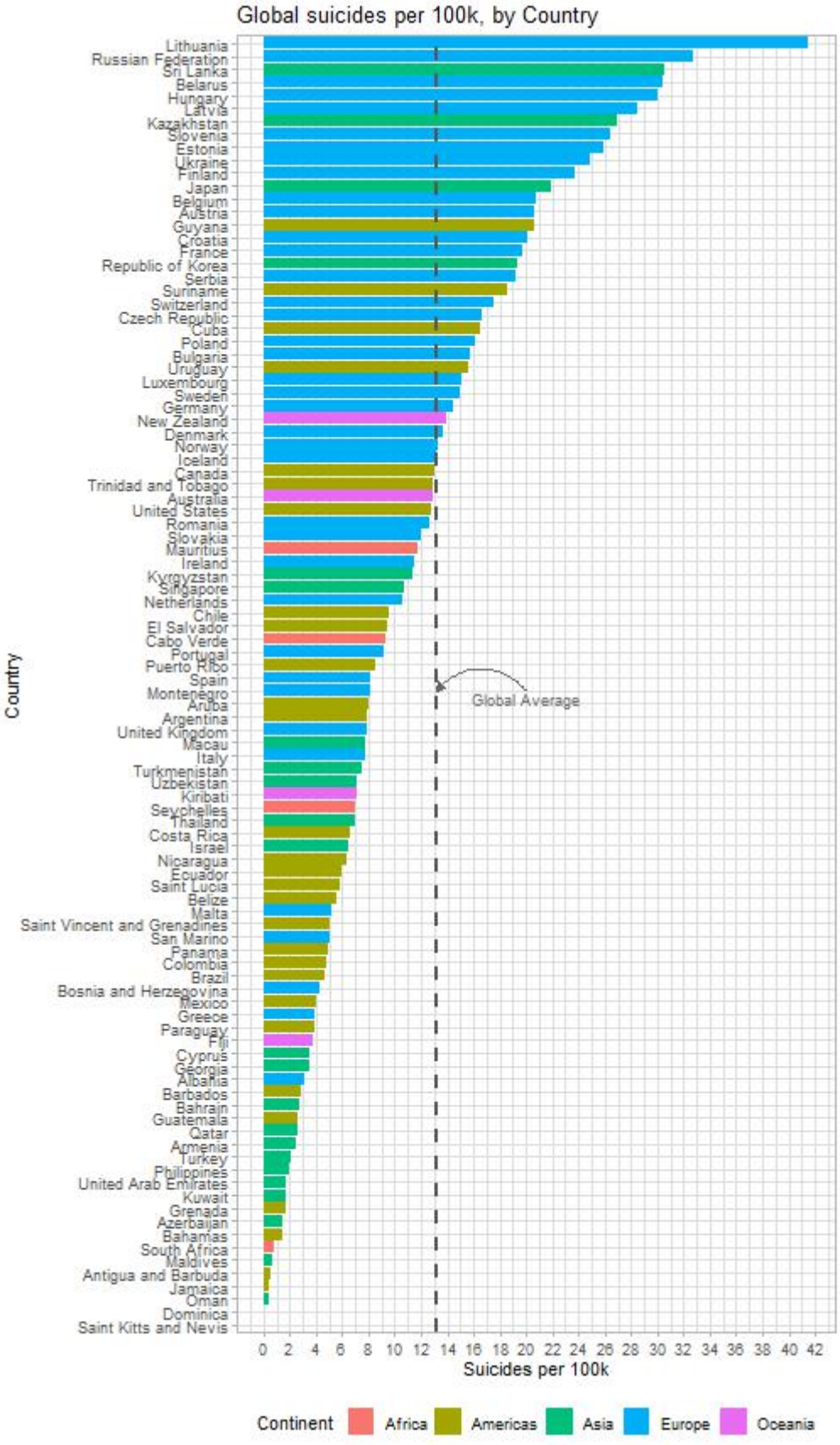
By Country

Overall

```
country <- data %>%
  group_by(country, continent) %>%
  summarize(n = n(),
            suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5) %>%
  arrange(desc(suicide_per_100k))
```

```
## `summarise()` has grouped output by 'country'. You can override using the
## `.groups` argument.
```

```
country$country <- factor(country$country,
                          ordered = T,
                          levels = rev(country$country))

ggplot(country, aes(x = country, y = suicide_per_100k, fill = continent)) +
  geom_bar(stat = "identity") +
  geom_hline(yintercept = global_average,
             linetype = 2, color = "grey35", size = 1) +
  annotate("text", x = 50, y = 20, label = "Global Average", vjust = 1, size = 3, color = "grey40") +
  annotate(
    "curve",
    x = 50, y = 20,
    xend = 50, yend = 13.1,
    arrow = arrow(length = unit(0.2, "cm"), type = "closed"),
    color = "grey40") +
  labs(title = "Global suicides per 100k, by Country",
           x = "Country",
           y = "Suicides per 100k",
        fill = "Continent") +
  coord_flip() +
  scale_y_continuous(breaks = seq(0, 45, 2)) +
  theme(legend.position = "bottom")
```

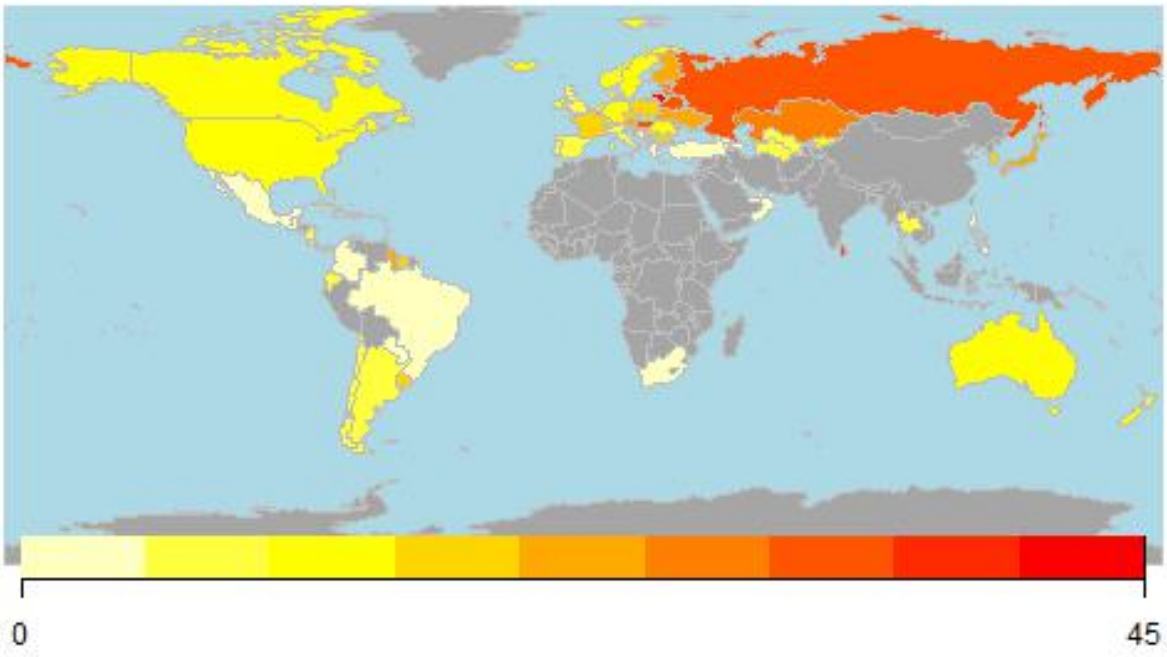

Global suicides per 100k, by Country
Lithuania
Russian Federation
Sri Lanka
Belarus
Hungary
Latvia
Kazakhstan
Slovenia
Estonia
Ukraine
Finland
Japan
Belgium
Austria
Guyana
Croatia
France
Republic of Korea
Serbia
Suriname
Switzerland
Czech Republic
Cuba
Poland
Bulgaria
Uruguay
Luxembourg
Sweden
Germany
New Zealand
Denmark
Norway
Iceland
Canada
Trinidad and Tobago
Australia
United States
Romania
Slovakia
Mauritius
Ireland
Kyrgyzstan
Singapore
Netherlands
Chile
El Salvador
Cabo Verde
Portugal
Puerto Rico
Spain
Montenegro
Aruba
Argentina
United Kingdom
Macau
Italy
Turkmenistan
Uzbekistan
Kiribati
Seychelles
Thailand
Costa Rica
Israel
Nicaragua
Ecuador
Saint Lucia
Belize
Malta
Saint Vincent and Grenadines
San Marino
Panama
Colombia
Brazil
Bosnia and Herzegovina
Mexico
Greece
Paraguay
Fiji
Cyprus
Georgia
Albania
Barbados
Bahrain
Guatemala
Qatar
Armenia
Turkey
Philippines
United Arab Emirates
Kuwait
Grenada
Azerbaijan
Bahamas
South Africa
Maldives
Antigua and Barbuda
Jamaica
Oman
Dominica
Saint Kitts and Nevis
Country
Global Average
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42
Suicides per 100k
Continent
Africa
Americas
Asia
Europe
Oceania

Insights

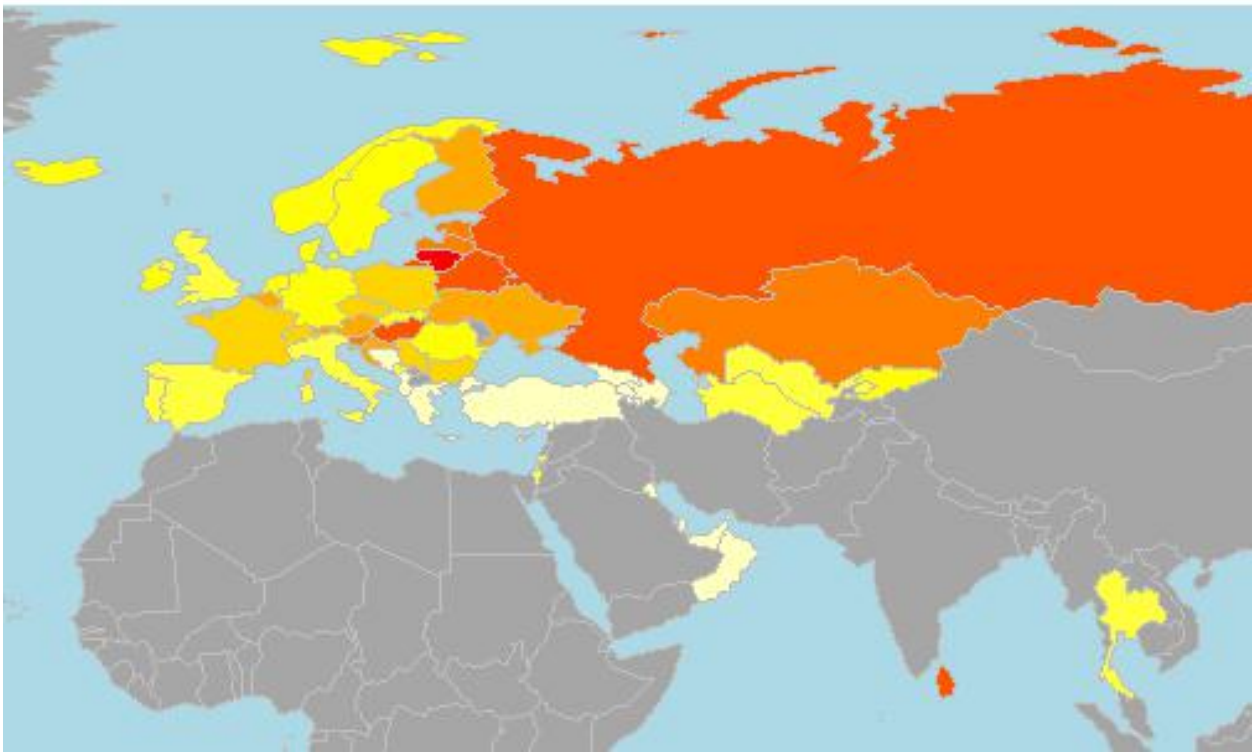
- Lithuania's rate has been highest by a large margin: > 41 suicides per 100k (per year)
- Large over representation of European countries with high rates, few with low rates

Below is a geographical heat map of the suicide rates between the timeframe of this analysis - **note the lack of data for Africa and Asia**, and bear in mind that 7 countries have been removed due to insufficient data.

```
country <- data %>%
  group_by(country) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5)

countrydata <- joinCountryData2Map(country, joinCode = "NAME", nameJoinColumn = "country")

par(mar=c(0, 0, 0, 0)) # margins

mapCountryData(countrydata,
nameColumnToPlot="suicide_per_100k",
mapTitle="",
colourPalette = "heat",
oceanCol="lightblue",
missingCountryCol="grey65",
catMethod = "pretty")
```



```
mapCountryData(countrydata,
nameColumnToPlot="suicide_per_100k",
mapTitle="",
mapRegion = "eurasia",
colourPalette = "heat",
oceanCol="lightblue",
missingCountryCol="grey65",
addLegend = FALSE,
catMethod = "pretty")
```



It's important to note that looking at figures at a global/continent level might not truly be representative of the globe/continent for these reasons.

Comparing the raw suicide rates of countries may also lead to some issues - the definition of suicide (and the reliability that a death is recorded as suicide) will likely vary between countries.

However, trends over time (within countries) are likely to be reliable. I address this next.

Linear Trends

I'm interested in how the suicide rate is changing over time within each country. Instead of visualizing all 93 countries rates across time, I fit a simple linear regression to every countries data. I extract those with a 'year' p-value (*corrected for multiple comparisons*) of < 0.05.

In other words: as time goes on, **I look for countries where the suicide rate is linearly increasing or decreasing over time**. These can then be rank ordered by their 'year' coefficient, which would be their rate of change as time goes on.

```
country_year <- data %>%
  group_by(country, year) %>%
  summarize(suicides         = sum(suicides_no),
            population       = sum(population),
            suicide_per_100k = (suicides / population) * 10^5,
            gdp_per_capita   = mean(gdp_per_capita))
```

```
## `summarise()` has grouped output by 'country'. You can override using the
## `.groups` argument.
```

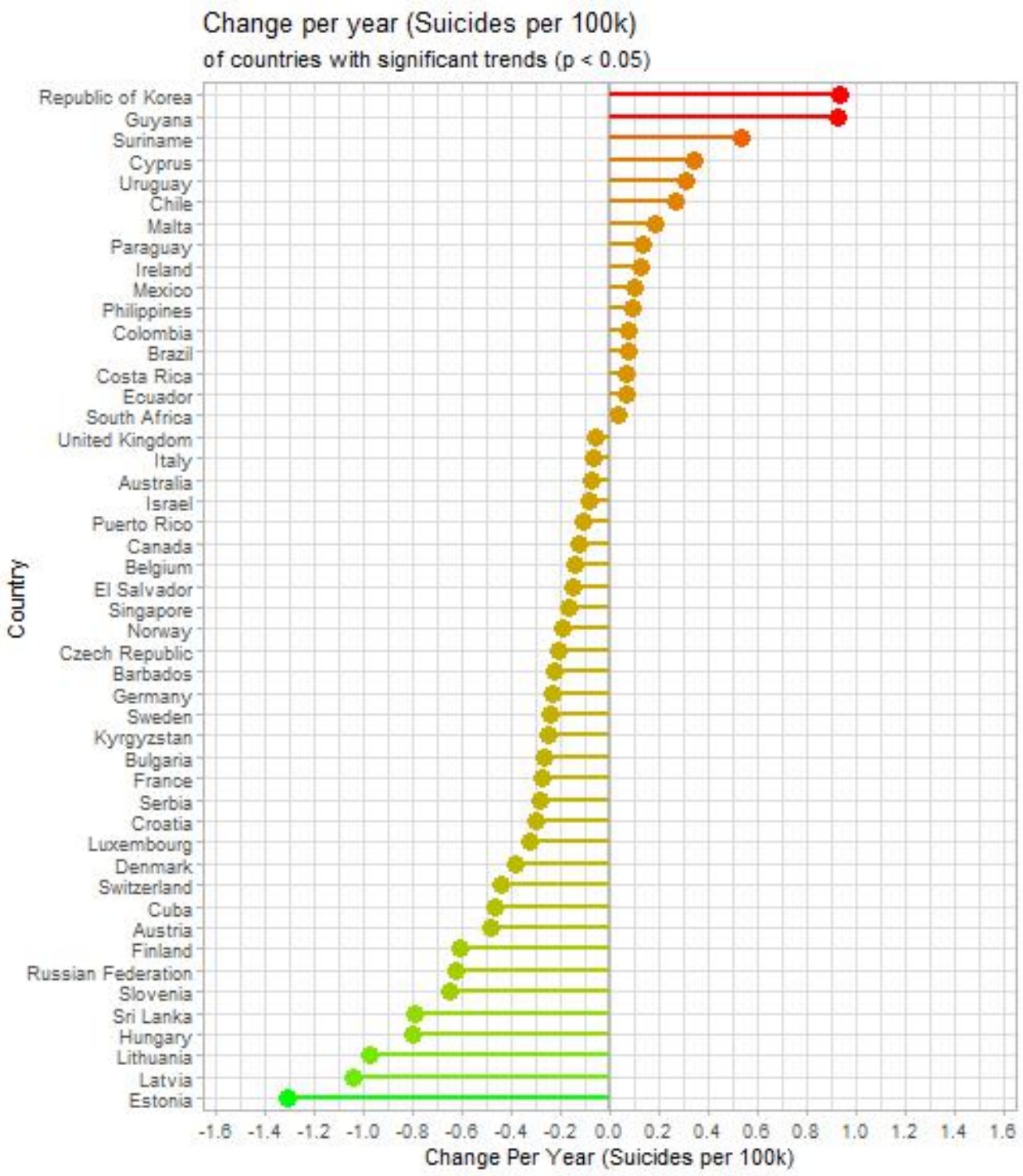
```
country_year_trends <- country_year %>%
  ungroup() %>%
  nest(-country) %>% # format: country, rest of data (in list column)
  mutate(model = map(data, ~ lm(suicide_per_100k ~ year, data = .)), # for each item in 'data', fit a l
         tidied = map(model, tidy)) %>% # tidy each of these into data frame format - call this list 't
  unnest(tidied)
```

```
## Warning: Supplying `...` without names was deprecated in tidyr 1.0.0.
## i Please specify a name for each selection.
## i Did you want `data = -country`?
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
country_year_sig_trends <- country_year_trends %>%
  filter(term == "year") %>%
  mutate(p.adjusted = p.adjust(p.value, method = "holm")) %>%
  filter(p.adjusted < .05) %>%
  arrange(estimate)

country_year_sig_trends$country <- factor(country_year_sig_trends$country,
                                          ordered = T,
                                          levels = country_year_sig_trends$country)
```

```
# plot 1
ggplot(country_year_sig_trends, aes(x = country, y = estimate, col = estimate)) +
  geom_point(stat = 'identity', size = 4) +
  geom_hline(yintercept = 0, col = "grey", size = 1) +
  scale_color_gradient(low = "green", high = "red") +
  geom_segment(aes(y    = 0,
                   x    = country,
                   yend = estimate,
                   xend = country), size = 1) +
  labs(title    = "Change per year (Suicides per 100k)",
```

```
      subtitle = "of countries with significant trends (p < 0.05)",
      x        = "Country",
      y        = "Change Per Year (Suicides per 100k)") +
scale_y_continuous(breaks = seq(-2, 2, 0.2), limits = c(-1.5, 1.5)) +
theme(legend.position = "none") +
coord_flip()
```



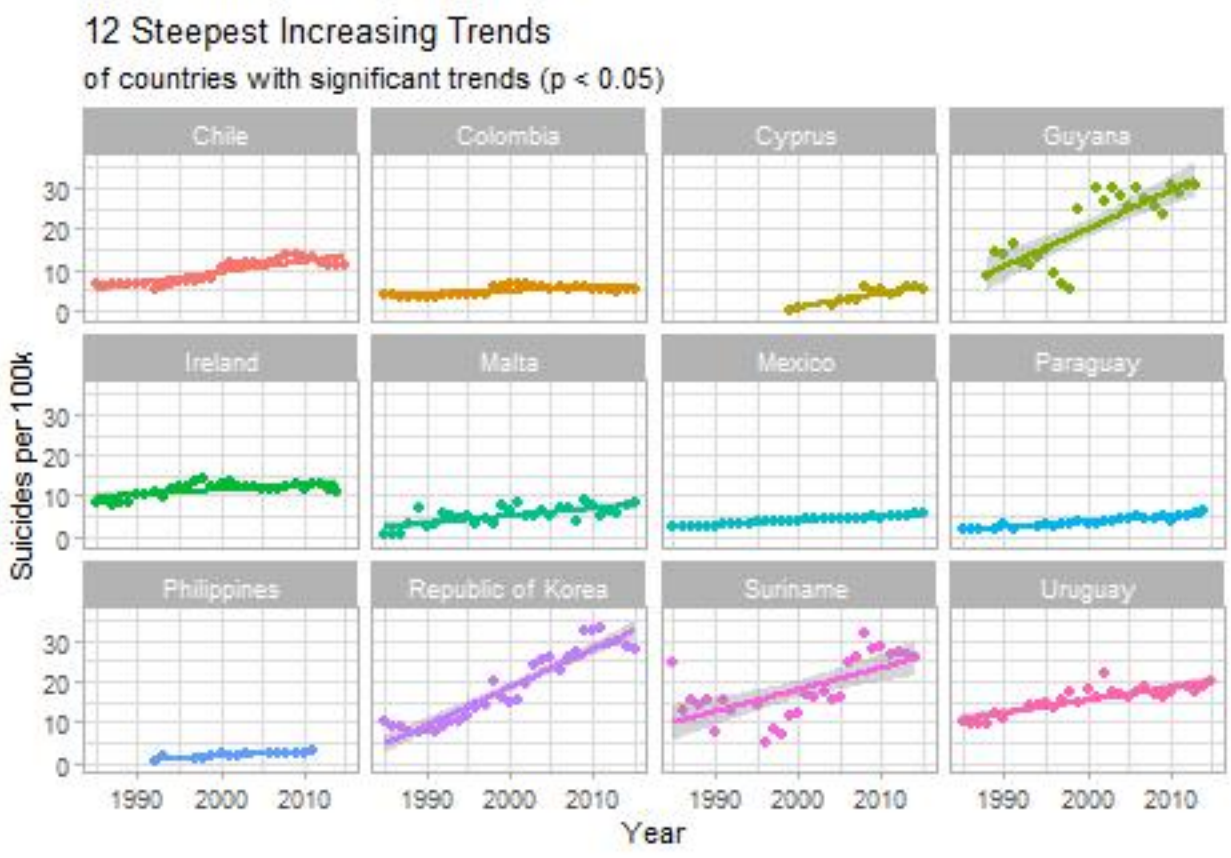
Insights

- ~**1/2** of all countries suicide rates are **changing linearly** as time progresses
- 32 (**2/3**) of these 48 countries are **decreasing**
- Overall, this is painting a positive picture

Steepest *increasing* trends:

```
### Lets look at those countries with the steepest increasing trends

top12_increasing <- tail(country_year_sig_trends$country, 12)

country_year %>%
  filter(country %in% top12_increasing) %>%
  ggplot(aes(x = year, y = suicide_per_100k, col = country)) +
  geom_point() +
  geom_smooth(method = "lm") +
  facet_wrap(~ country) +
  theme(legend.position = "none") +
  labs(title = "12 Steepest Increasing Trends",
       subtitle = "of countries with significant trends (p < 0.05)",
       x = "Year",
       y = "Suicides per 100k")
```

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



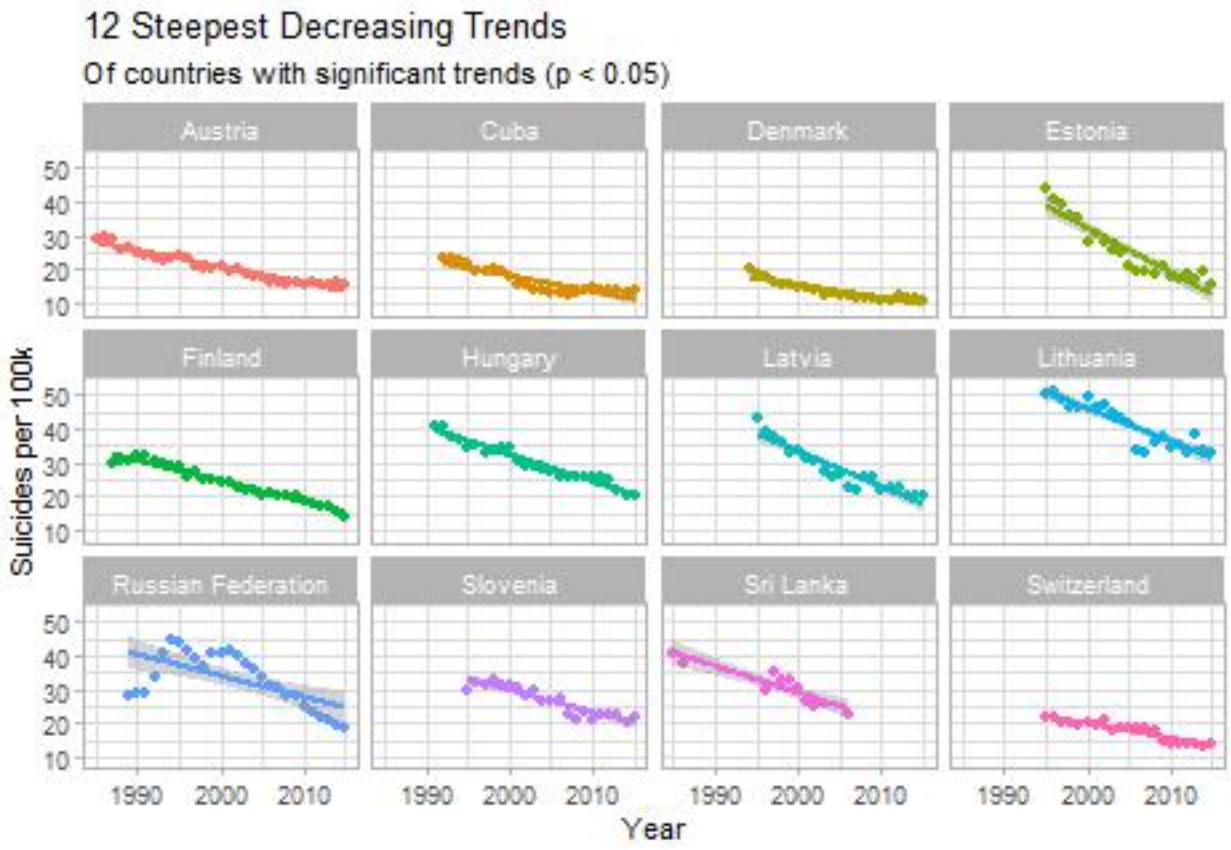
Insights

- **South Korea** shows the most concerning trend - an increase in suicide of 0.931 people (per 100k, per year) - the **steepest increase globally**
- **Guyana** is similar, at + 0.925 people (per 100k, per year)
- Between **1998 and 1999 (5.3 to 24.8)**, Guyana's rate increased by ~365%
- The historical data for Guyana seems questionable - it's known for very high suicide rates but the jump seems unlikely (maybe changed how they classified suicide?)

Steepest *decreasing* trends:

```
### Now those with the steepest decreasing trend

top12_decreasing <- head(country_year_sig_trends$country, 12)

country_year %>%
  filter(country %in% top12_decreasing) %>%
  ggplot(aes(x = year, y = suicide_per_100k, col = country)) +
  geom_point() +
  geom_smooth(method = "lm") +
  facet_wrap(~ country) +
  theme(legend.position = "none") +
  labs(title    = "12 Steepest Decreasing Trends",
       subtitle = "Of countries with significant trends (p < 0.05)",
       x = "Year",
       y = "Suicides per 100k")
```

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



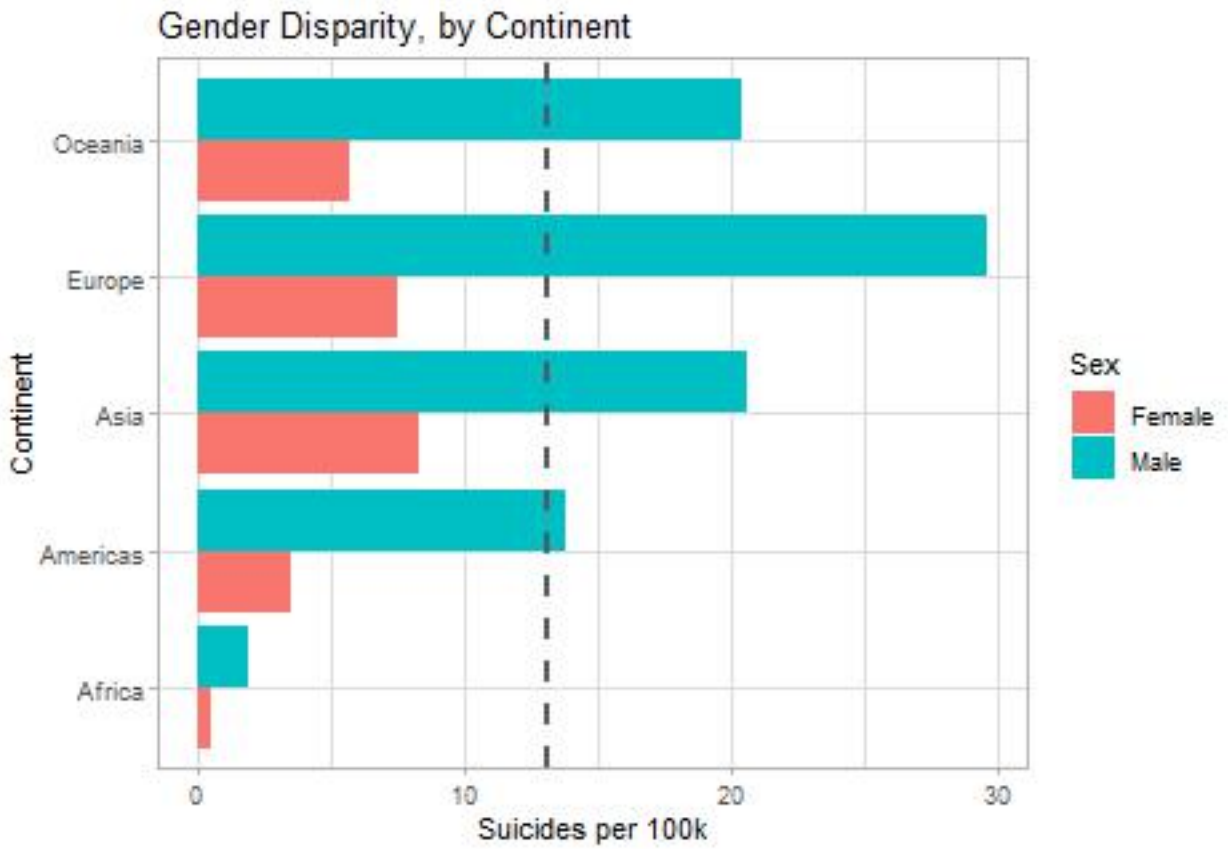
Insights

- **Estonia** shows the most positive trend - **every year, ~1.31 *less* people (per 100k) commit suicide** - the steepest decrease globally
- Between 1995 and 2015, this drops from **43.8 to 15.7** per 100k (per year) - a **64% decrease**
- The Russian Federation trend is interesting, only beginning to drop in 2002. Since then it has decreased by ~50%.

Gender differences, by Continent

```
data %>%
  group_by(continent, sex) %>%
  summarize(n = n(),
            suicides    = sum(as.numeric(suicides_no)),
            population = sum(as.numeric(population)),
            suicide_per_100k = (suicides / population) * 100000) %>%
  ggplot(aes(x = continent, y = suicide_per_100k, fill = sex)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_hline(yintercept = global_average, linetype = 2, color = "grey35", size = 1) +
  labs(title = "Gender Disparity, by Continent",
   x = "Continent",
   y = "Suicides per 100k",
   fill = "Sex") +
  coord_flip()
```

```
## `summarise()` has grouped output by 'continent'. You can override using the
## `.groups` argument.
```



Insights

- **European men** were at the **highest risk** between 1985 - 2015, at ~ 30 suicides (per 100k, per year)
- **Asia** had the smallest overrepresentation of male suicide - the rate was ~**2.5x as high for men**
- Comparatively, **Europe's** rate was ~**3.9x as high for men**

Gender differences, by Country

```
country_long <- data %>%
  group_by(country, continent) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5) %>%
  mutate(sex = "OVERALL")
```

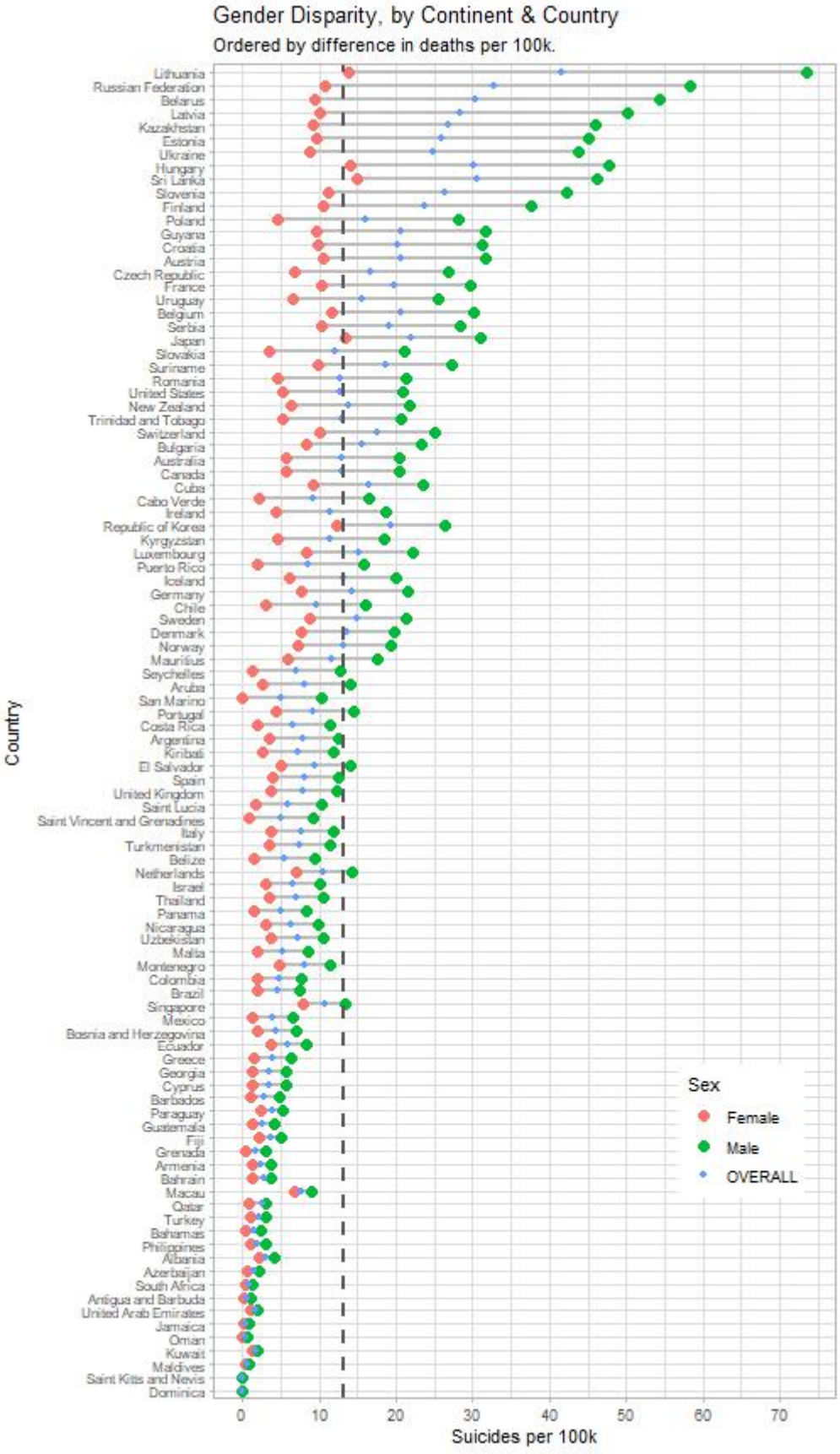
```
## `summarise()` has grouped output by 'country'. You can override using the
## `.groups` argument.
```

```
### by country, continent, sex

sex_country_long <- data %>%
  group_by(country, continent, sex) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5)
```

```
## `summarise()` has grouped output by 'country', 'continent'. You can override
## using the `.groups` argument.
```

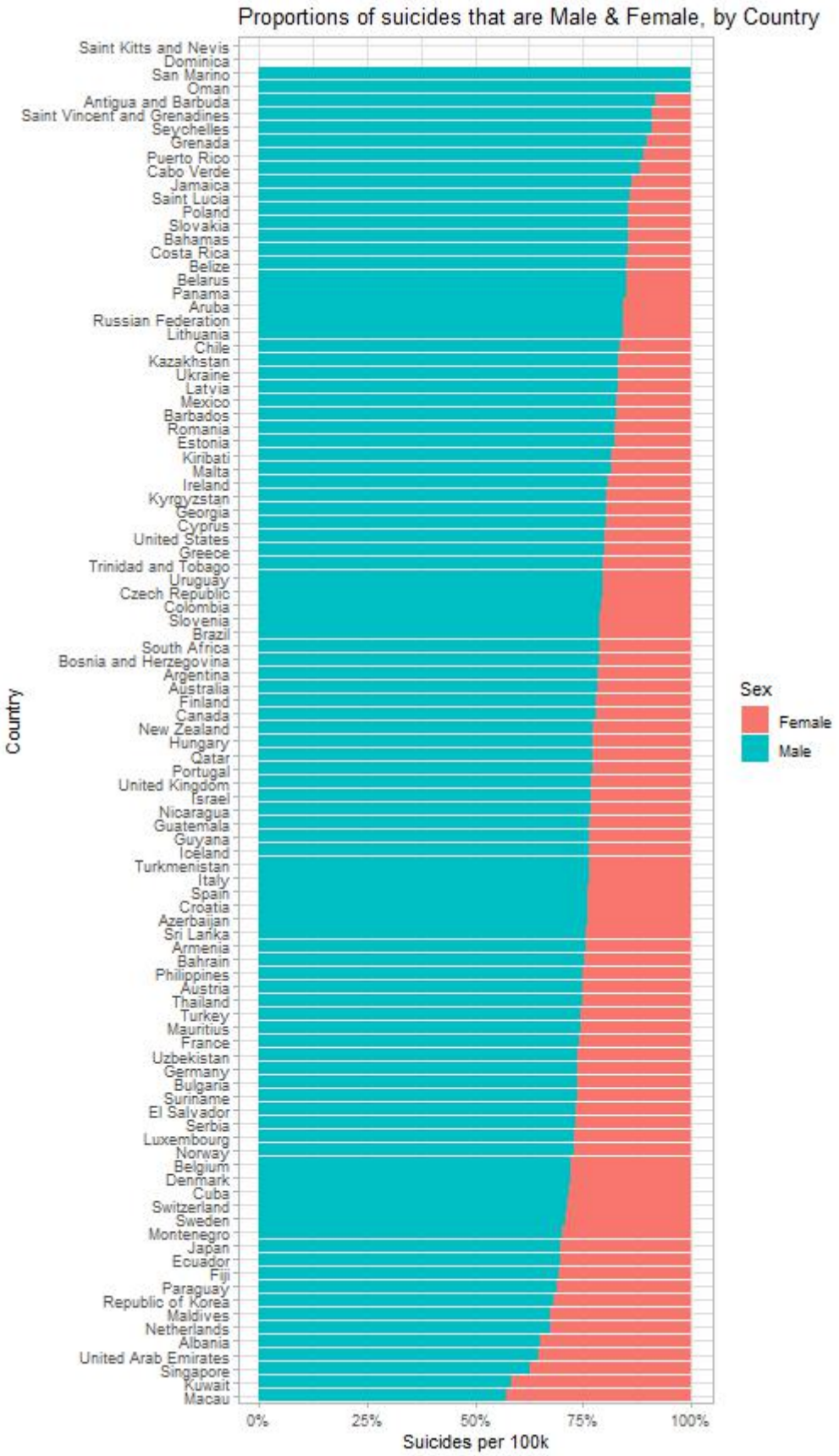
```
sex_country_wide <- sex_country_long %>%
  spread(sex, suicide_per_100k) %>%
  arrange(Male - Female)


sex_country_wide$country <- factor(sex_country_wide$country,
                                   ordered = T,
                                   levels = sex_country_wide$country)

sex_country_long$country <- factor(sex_country_long$country,
                                   ordered = T,
                                   levels = sex_country_wide$country) # using the same order



### this graph shows us how the disparity between deaths varies across gender for every country
# it also has the overall blended death rate - generally countries with a higher death rate have a high
# this is because, if suicide is more likely in a country, the disparity between men and women is ampli

ggplot(sex_country_wide, aes(y = country, color = sex)) +
  geom_dumbbell(aes(x=Female, xend=Male), color = "grey", size = 1) +
  geom_point(data = sex_country_long, aes(x = suicide_per_100k), size = 3) +
  geom_point(data = country_long, aes(x = suicide_per_100k)) +
  geom_vline(xintercept = global_average, linetype = 2, color = "grey35", size = 1) +
  theme(axis.text.y = element_text(size = 8),
        legend.position = c(0.85, 0.2)) +
  scale_x_continuous(breaks = seq(0, 80, 10)) +
  labs(title    = "Gender Disparity, by Continent & Country",
       subtitle = "Ordered by difference in deaths per 100k.",
       x = "Suicides per 100k",
       y = "Country",
       color = "Sex")
```

```
## Warning: A numeric ‘legend.position‘ argument in ‘theme()‘ was deprecated in ggplot2
## 3.5.0.
## i Please use the ‘legend.position.inside‘ argument of ‘theme()‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.

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


Gender Disparity, by Continent & Country
Ordered by difference in deaths per 100k.
Lithuania
Russian Federation
Belarus
Latvia
Kazakhstan
Estonia
Ukraine
Hungary
Sri Lanka
Slovenia
Finland
Poland
Guyana
Croatia
Austria
Czech Republic
France
Uruguay
Belgium
Serbia
Japan
Slovakia
Suriname
Romania
United States
New Zealand
Trinidad and Tobago
Switzerland
Bulgaria
Australia
Canada
Cuba
Cabo Verde
Ireland
Republic of Korea
Kyrgyzstan
Luxembourg
Puerto Rico
Iceland
Germany
Chile
Sweden
Denmark
Norway
Mauritius
Seychelles
Aruba
San Marino
Portugal
Costa Rica
Argentina
Kiribati
El Salvador
Spain
United Kingdom
Saint Lucia
Saint Vincent and Grenadines
Italy
Turkmenistan
Belize
Netherlands
Israel
Thailand
Panama
Nicaragua
Uzbekistan
Malta
Montenegro
Colombia
Brazil
Singapore
Mexico
Bosnia and Herzegovina
Ecuador
Greece
Georgia
Cyprus
Barbados
Paraguay
Guatemala
Fiji
Grenada
Armenia
Bahrain
Macau
Qatar
Turkey
Bahamas
Philippines
Albania
Azerbaijan
South Africa
Antigua and Barbuda
United Arab Emirates
Jamaica
Oman
Kuwait
Maldives
Saint Kitts and Nevis
Dominica
Country
0
10
20
30
40
50
60
70
Suicides per 100k
Sex
Female
Male
OVERALL

```r
country_gender_prop <- sex_country_wide %>%
  mutate(Male_Proportion = Male / (Female + Male)) %>%
  arrange(Male_Proportion)

sex_country_long$country <- factor(sex_country_long$country,
                                   ordered = T,
                                   levels = country_gender_prop$country)

ggplot(sex_country_long, aes(y = suicide_per_100k, x = country, fill = sex)) +
  geom_bar(position = "fill", stat = "identity") +
  scale_y_continuous(labels = scales::percent) +
  labs(title = "Proportions of suicides that are Male & Female, by Country",
       x = "Country",
       y = "Suicides per 100k",
       fill = "Sex") +
  coord_flip()
```

```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_bar()`).
```


Proportions of suicides that are Male & Female, by Country
Saint Kitts and Nevis
Dominica
San Marino
Oman
Antigua and Barbuda
Saint Vincent and Grenadines
Seychelles
Grenada
Puerto Rico
Cabo Verde
Jamaica
Saint Lucia
Poland
Slovakia
Bahamas
Costa Rica
Belize
Belarus
Panama
Aruba
Russian Federation
Lithuania
Chile
Kazakhstan
Ukraine
Latvia
Mexico
Barbados
Romania
Estonia
Kiribati
Malta
Ireland
Kyrgyzstan
Georgia
Cyprus
United States
Greece
Trinidad and Tobago
Uruguay
Czech Republic
Colombia
Slovenia
Brazil
South Africa
Bosnia and Herzegovina
Argentina
Australia
Finland
Canada
New Zealand
Hungary
Qatar
Portugal
United Kingdom
Israel
Nicaragua
Guatemala
Guyana
Iceland
Turkmenistan
Italy
Spain
Croatia
Azerbaijan
Sri Lanka
Armenia
Bahrain
Philippines
Austria
Thailand
Turkey
Mauritius
France
Uzbekistan
Germany
Bulgaria
Suriname
El Salvador
Serbia
Luxembourg
Norway
Belgium
Denmark
Cuba
Switzerland
Sweden
Montenegro
Japan
Ecuador
Fiji
Paraguay
Republic of Korea
Maldives
Netherlands
Albania
United Arab Emirates
Singapore
Kuwait
Macau
Country
0%
25%
50%
75%
100%
Suicides per 100k
Sex
Female
Male

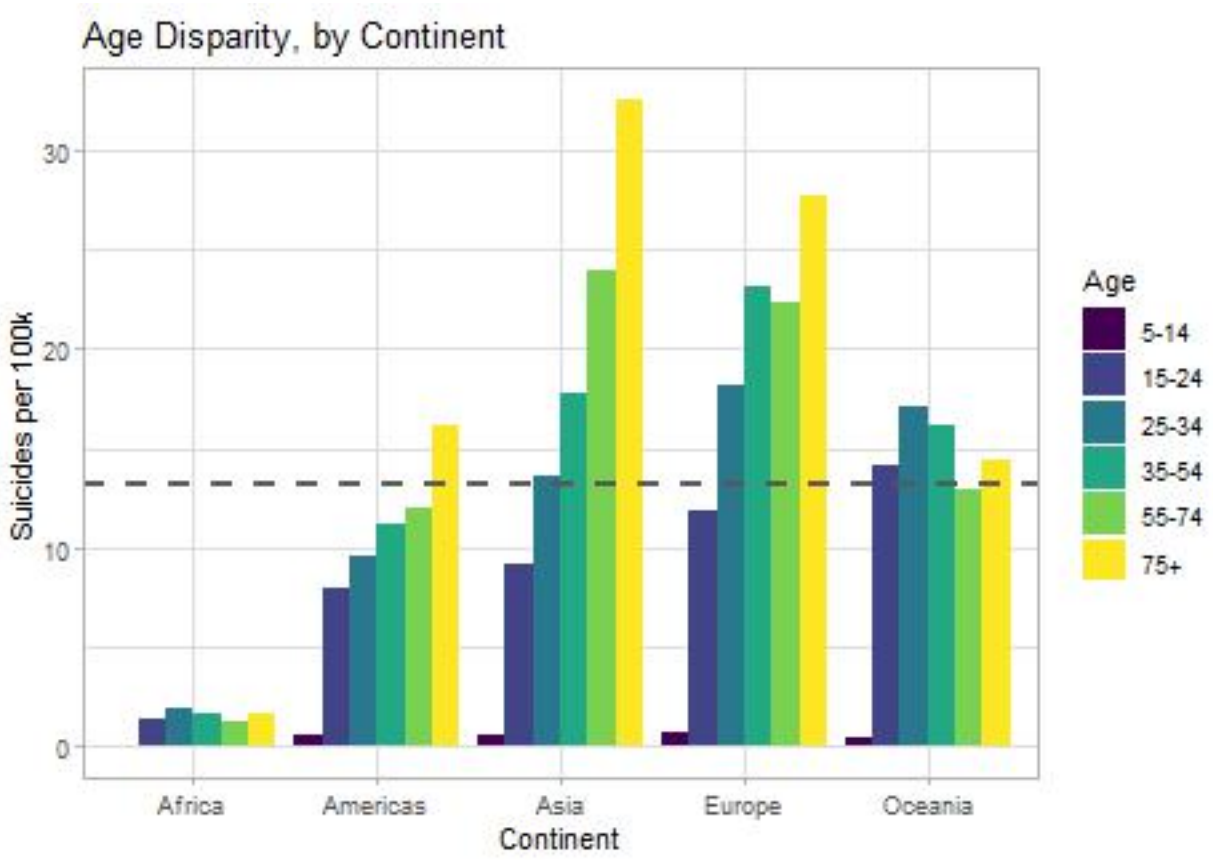
Insights

- The **overrepresentation of men** in suicide deaths appears to be **universal**, and can be observed to differing extents in every country
- Whilst **women are more likely to suffer from depression and suicidal thoughts, men are more likely to die from suicide**
- This is known as the gender paradox on suicidal behaviour

Age differences, by Continent

```
data %>%
  group_by(continent, age) %>%
  summarize(n = n(),
            suicides   = sum(as.numeric(suicides_no)),
            population = sum(as.numeric(population)),
            suicide_per_100k = (suicides / population) * 10^5) %>%
  ggplot(aes(x = continent, y = suicide_per_100k, fill = age)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_hline(yintercept = global_average, linetype = 2, color = "grey35", size = 1) +
  labs(title = "Age Disparity, by Continent",
       x = "Continent",
       y = "Suicides per 100k",
       fill = "Age")
```

```
## `summarise()` has grouped output by 'continent'. You can override using the
## `.groups` argument.
```



Insights

- For the **Americas, Asia & Europe** (which make up most of the dataset), **suicide rate increases with age**
- **Oceania & Africa's rates are highest for those aged 25 - 34**

As a country gets richer, does it's suicide rate decrease?

GDP maybe a variable that may concern by people who live in that particular country. Maybe it's the one factor that cause suicide rate to increase, that why I want to perform some test to see what really going on here. I'm asking for the cause in this section, but I'm not finding the cause in this section, I'm looking for the association here.

It depends on the country - for almost every country, there is a high correlation between `year` and `gdp per capita`, i.e. as time goes on, gdp per capita linearly increases.

```
country_year_gdp <- data %>%
  group_by(country, year) %>%
  summarize(gdp_per_capita = mean(gdp_per_capita))
```

```
## `summarise()` has grouped output by 'country'. You can override using the
## `.groups` argument.
```

```
country_year_gdp_corr <- country_year_gdp %>%
  ungroup() %>%
  group_by(country) %>%
  summarize(year_gdp_correlation = cor(year, gdp_per_capita))
```

I calculated the pearson correlations between 'year' and 'GDP per capita' within each country, then summarized the results:

The mean correlation was **NA**, indicating a *very* strong positive linear relationship.

This basically means that looking within a country and asking **"does an increase in weath (per person) have an effect suicide rate"** is pretty similar to asking **"does a countries suicide rate increase as time progresses"**.

This was answered earlier in **(2.5.2)** - it depends on the country! Some countries are increasing with time, most are decreasing.

Instead, I ask a slightly different question below.

Do richer countries have a higher rate of suicide?

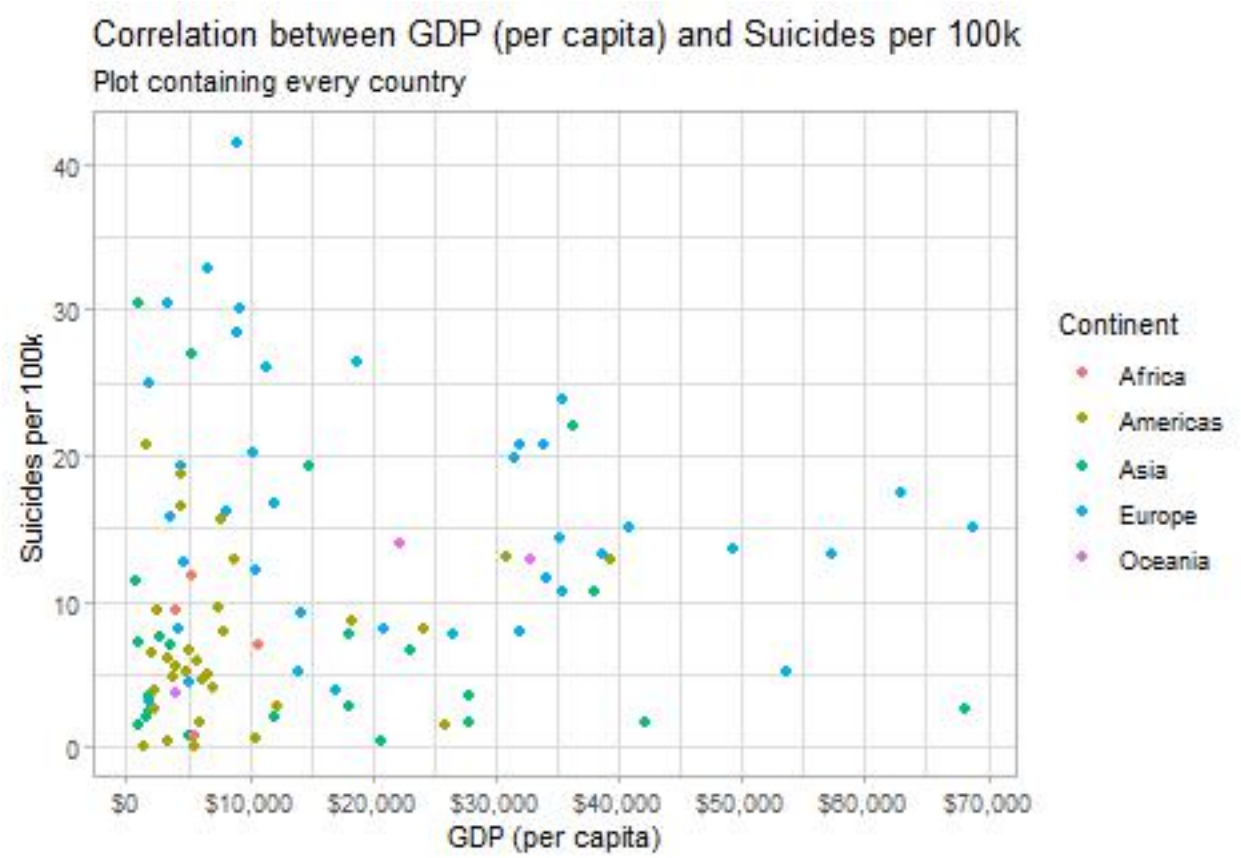
Instead of looking at trends within countries, here I take every country and calculate their mean GDP (per capita) across all the years in which data is available. I then measure how this relates to the countries suicide rate across all those years.

The end result is one data point per country, intended to give a general idea of the wealth of a country and its suicide rate.

```
country_mean_gdp <- data %>%
  group_by(country, continent) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) / sum(as.numeric(population))) * 100000,
            gdp_per_capita = mean(gdp_per_capita))
```

```
## `summarise()` has grouped output by 'country'. You can override using the
## `.groups` argument.
```

```
ggplot(country_mean_gdp, aes(x = gdp_per_capita, y = suicide_per_100k, col = continent)) +
  geom_point() +
  scale_x_continuous(labels=scales::dollar_format(prefix="$"), breaks = seq(0, 70000, 10000)) +
  labs(title = "Correlation between GDP (per capita) and Suicides per 100k",
       subtitle = "Plot containing every country",
       x = "GDP (per capita)",
       y = "Suicides per 100k",
       col = "Continent")
```



There are quite a few high leverage & residual countries that could have a significant impact on the fit of my regression line (e.g. Lithuania, top left). I'll identify and exclude these using Cooks Distance, excluding those countries with a CooksD value of greater than 4/n.

I assess the statistics of this model (with outliers removed) below.

```
model1 <- lm(suicide_per_100k ~ gdp_per_capita, data = country_mean_gdp)

gdp_suicide_no_outliers <- model1 %>%
  augment() %>%
  arrange(desc(.cooksd)) %>%
  filter(.cooksd < 4/nrow(.)) %>% # removes 5/93 countries
  inner_join(country_mean_gdp, by = c("suicide_per_100k", "gdp_per_capita")) %>%
  select(country, continent, gdp_per_capita, suicide_per_100k)

model2 <- lm(suicide_per_100k ~ gdp_per_capita, data = gdp_suicide_no_outliers)

summary(model2)
```

```
##
## Call:
## lm(formula = suicide_per_100k ~ gdp_per_capita, data = gdp_suicide_no_outliers)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.118  -5.100  -1.493   3.069  20.671
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    8.378e+00  1.073e+00   7.807 8.58e-12 ***
## gdp_per_capita 1.054e-04  4.798e-05   2.197   0.0305 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.339 on 93 degrees of freedom
## Multiple R-squared:  0.04934,    Adjusted R-squared:  0.03912
## F-statistic: 4.827 on 1 and 93 DF,  p-value: 0.03051
```

The **p-value** of the model is **0.0305** < 0.05. This means we can *reject* the hypothesis that a countries GDP (per capita) has no association with it's rate of suicide (per 100k).

The r-squared is **0.0493**, so GDP (per capita) explains very little of the variance in suicide rate overall.

What does all this mean?

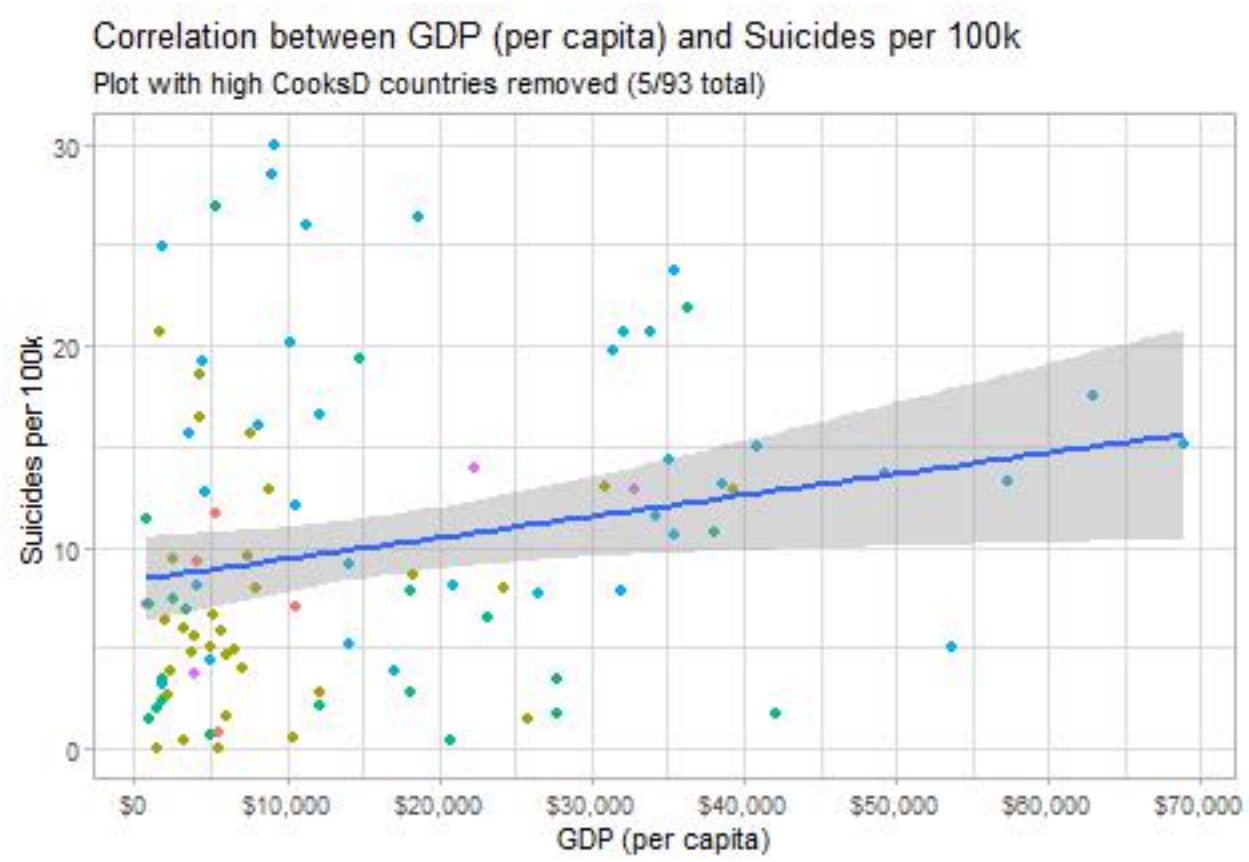
There is a weak but significant positive linear relationship - ***richer* countries are associated with *higher* rates of suicide**, but this is a ***weak* relationship** which can be seen from the graph below.

```
ggplot(gdp_suicide_no_outliers, aes(x = gdp_per_capita, y = suicide_per_100k, col = continent)) +
  geom_point() +
  geom_smooth(method = "lm", aes(group = 1)) +
  scale_x_continuous(labels=scales::dollar_format(prefix="$"), breaks = seq(0, 70000, 10000)) +
  labs(title = "Correlation between GDP (per capita) and Suicides per 100k",
       subtitle = "Plot with high CooksD countries removed (5/93 total)",
       x = "GDP (per capita)",
```

```
       y = "Suicides per 100k",
       col = "Continent") +
  theme(legend.position = "none")
```

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

```
## Warning: The following aesthetics were dropped during statistical transformation:
## colour.
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



This line of best fit is represented by the equation below, where:

- Suicides = Suicides per 100k
- GDP = GDP per capita (in thousands, USD)

$$Suicides = 8.7718 + 0.1115 * GDP$$

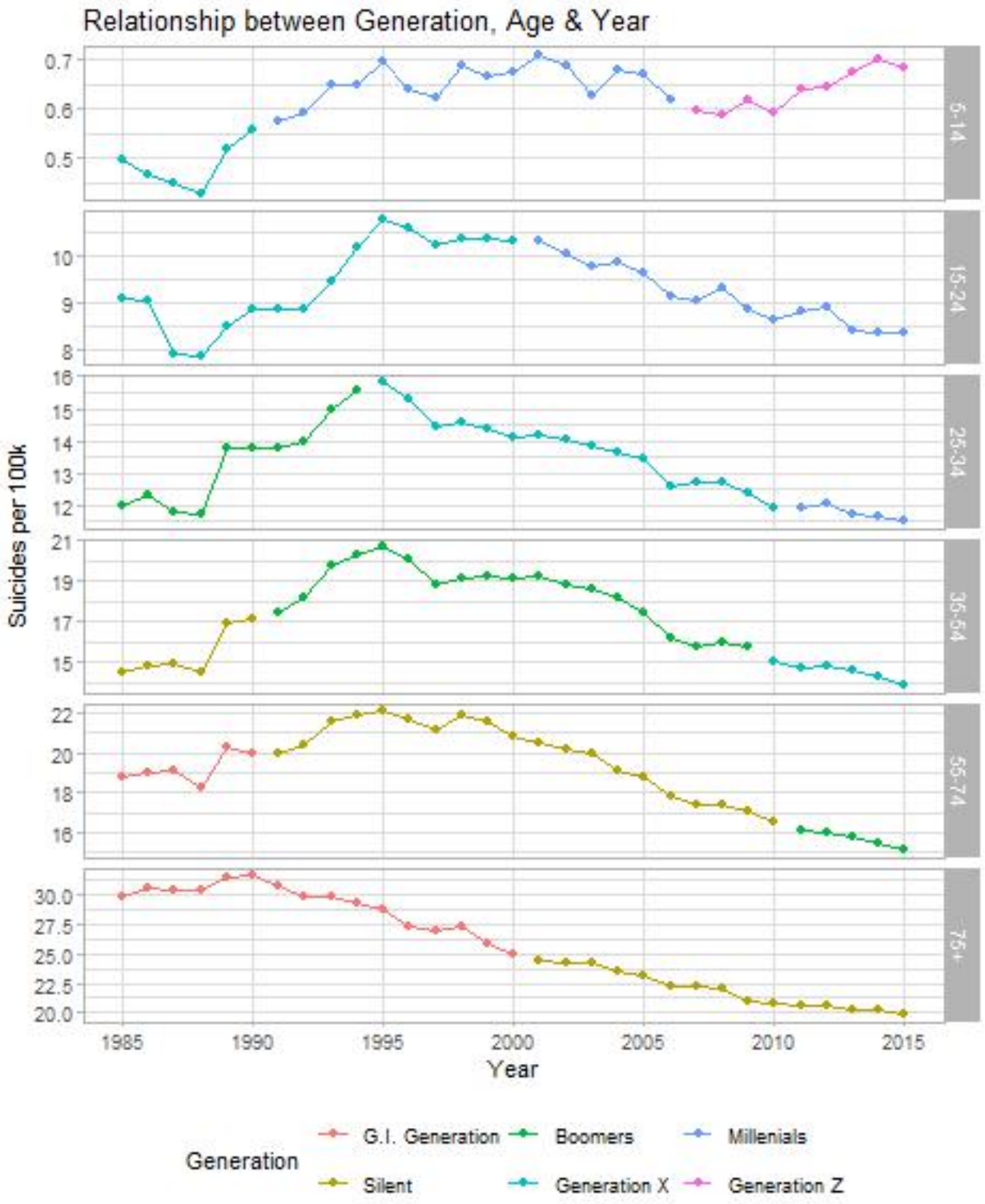
This means that, at a country level and over the time frame of this analysis (1985 - 2015), **an increase of GDP (per capita) by $8,967 was associated with 1 additional suicide, per 100k people, per year**.

The reason I haven't used the generation variable

With *continuous data*, if you have someones age in a given year, you have their generation. The graph below demonstrates how this works for this dataset really well, and is equivalent to the graph of age across time, shown in **(2.4)**.

```
data %>%
  group_by(generation, age, year) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) / sum(as.numeric(population))) * 10^5) %>%
  ggplot(aes(x = year, y = suicide_per_100k,
             col = factor(generation, ordered = F))) +
  geom_point() +
  geom_line() +
  facet_grid(age ~ ., scales = "free_y") +
  scale_x_continuous(breaks = seq(1985, 2015, 5), minor_breaks = NULL) +
  labs(title = "Relationship between Generation, Age & Year",
       x = "Year",
       y = "Suicides per 100k",
       col = "Generation") +
  theme(legend.position = "bottom")
```

```
## `summarise()` has grouped output by 'generation', 'age'. You can override using
## the `.groups` argument.
```



However, because of the overlap of different age categories, trying to interpret the trend of generation suicide rates over time creates problems.

Compare the rates below to the plot above - **large spikes** occur at the same time that different age bands begin/stop being classified as from a certain generation.

Note, for example, the **supposed spike in suicide rate for G.I. generation in 1991**, where those aged '55 - 75' suddenly stop being classified as from this generation.

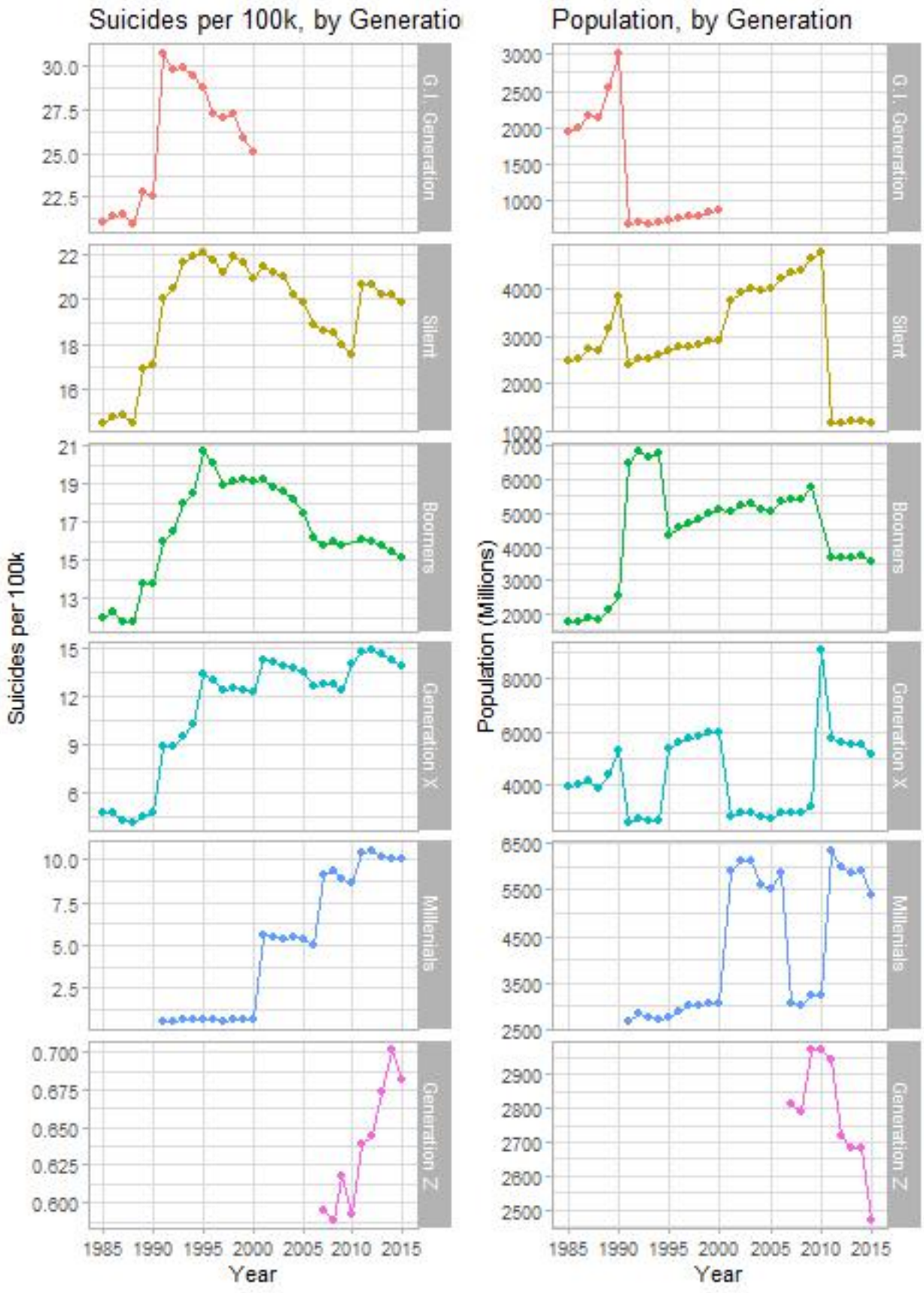
```
generation_rate <- data %>%
  group_by(generation, year) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5) %>%
  ggplot(aes(x = year, y = suicide_per_100k,
             col = factor(generation, ordered = F))) +
  geom_point() +
  geom_line() +
  facet_grid(generation ~ ., scales = "free_y") +
  scale_x_continuous(breaks = seq(1985, 2015, 5), minor_breaks = NULL) +
  labs(title = "Suicides per 100k, by Generation",
       x = "Year",
       y = "Suicides per 100k") +
  theme(legend.position = "none")
```

```
## `summarise()` has grouped output by 'generation'. You can override using the
## `.groups` argument.
```

```
generation_population <- data %>%
  group_by(generation, year) %>%
  summarize(population = sum(as.numeric(population))) %>%
  ggplot(aes(x = year, y = population / 10^5,
             col = factor(generation, ordered = F))) +
  geom_point() +
  geom_line() +
  facet_grid(generation ~ ., scales = "free_y") +
  scale_x_continuous(breaks = seq(1985, 2015, 5), minor_breaks = NULL) +
  labs(title = "Population, by Generation",
       x     = "Year",
       y     = "Population (Millions)",
       col   = "Generation") +
  theme(legend.position = "none")
```

```
## `summarise()` has grouped output by 'generation'. You can override using the
## `.groups` argument.
```

```
grid.arrange(generation_rate, generation_population, ncol = 2)
```


Suicides per 100k, by Generatio
Population, by Generation
Suicides per 100k
Population (Millions)
G.I. Generation
Silent
Boomers
Generation X
Millenials
Generation Z
Year
Year

This is probably a **problem with how the dataset was created** - it looks like the generation variable was created after the data was summarized (by country, year, age, sex) and just appended onto the end. This *shouldn't be possible*, because not everyone in a given age band & year will be of one generation.

This shows why the 'spikes' in generation across time are pretty meaningless and I would **recommend to others not to use the variable**, as it can will lead to wrong conclusions.

The 5% highest risk instances in history

I'll filter out data from 1985 only and look at what happens in the 3 decades following.

Here i'm interested in the 5% highest risk (suicides/100k) demographics between 1986 and 2015.

I define a *demographic* as a **year** in a particular **country**, for some combination of **sex** & **age**. e.g. 'United Kingdom, 2010, Female, 15 - 24' would be a single demographic/point on the jitter plot below.

In order for a demographic to be in the top 5% for historic suicide rates, it would require a suicide rate exceeding 50.7 (per 100k) in that year.

```
demographic_most <- data %>%
  mutate(suicides_per_100k = suicides_no * 10^5 / population) %>%
  arrange(desc(suicides_per_100k)) %>%
  filter(year != 1985) %>%
  head(n = round(nrow(.) * 5 / 100))


demographic_most$time <- ifelse(demographic_most$year <= 1995, "1986 - 1995",
                                ifelse(demographic_most$year <= 2005, "1996 - 2005",
                                       "2006 - 2015"))
```

```
ggplot(demographic_most, aes(x = age, fill = sex)) +
  geom_bar() +
  labs(title = "5% Most At-Risk Instances in History",
       subtitle = "Volumes by Decade, Age & Sex",
       x        = "Age",
       y        = "Number of Demographics",
       fill = "Sex") +
  facet_wrap(~ time) +
  scale_y_continuous(breaks = seq(0, 300, 20))
```


```
set.seed(1)

ggplot(demographic_most, aes(x = age, y = suicides_per_100k, col = sex)) +
  geom_jitter(alpha = 0.5) +
  labs(title = "5% Most At-Risk Instances in History",
       subtitle = "Instances by Decade, Age, & Sex",
       x        = "Age",
       y        = "Suicides per 100k",
       col = "Sex") +
  facet_wrap(~ time) +
  scale_y_continuous(breaks = seq(50, 300, 10))
```

Insights

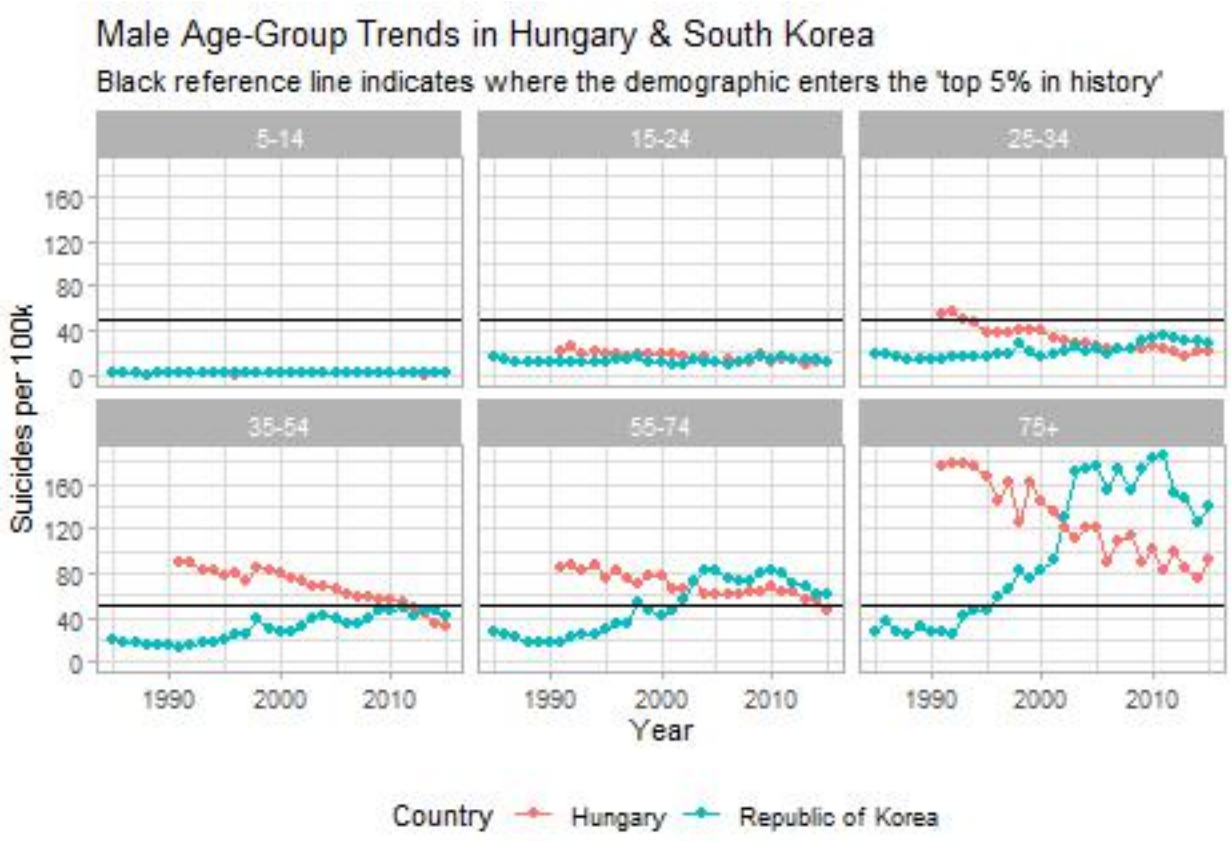
- **44.5%** of these 'high risk' instances occurred **between 1996 and 2005**
- **53.5%** were in the **75+** age category
- **96.9% were a male demographic**
- Of the 3.1% (42 instances) that *were* for **women**, **41/42** of these were in the **75+** demographic
- The **highest suicide rate** for a demographic in any year is **225 (per 100k)** - that's 0.225% of the entire demographic committing suicide in 1 year

Two of the most **consistently at-risk** demographics seem to be **men** in **South Korea & Hungary**, which I will visualize below.

```
data %>%
  filter(country %in% c('Republic of Korea', 'Hungary'), sex == "Male") %>%
  group_by(country, age, year) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) / sum(as.numeric(population))) * 100000) %
  ggplot(aes(x = year, y = suicide_per_100k, col = country)) +
  geom_line() +
  geom_point() +
  facet_wrap(~ age) +
  geom_hline(yintercept = min(demographic_most$suicides_per_100k)) +
  theme(legend.position = "bottom") +
  scale_y_continuous(breaks = seq(0, 220, 40)) +
  labs(title = "Male Age-Group Trends in Hungary & South Korea",
       subtitle = "Black reference line indicates where the demographic enters the 'top 5% in history'"
       x = "Year",
```

```
       y = "Suicides per 100k",
       col = "Country")
```

```
## `summarise()` has grouped output by 'country', 'age'. You can override using
## the `.groups` argument.
```



Two very different trends emerge. Hungary is obviously moving in a positive direction, whereas South Korea appears to be coming out of somewhat of a crisis.

For **South Korea**, **mens rates in the 75+ category** increased from 26.2 (per 100k) in 1992, to a peak of 185 (per 100k) in 2011 - **an *increase* of more than 600%**. Men aged 55-74 see a similar increase.

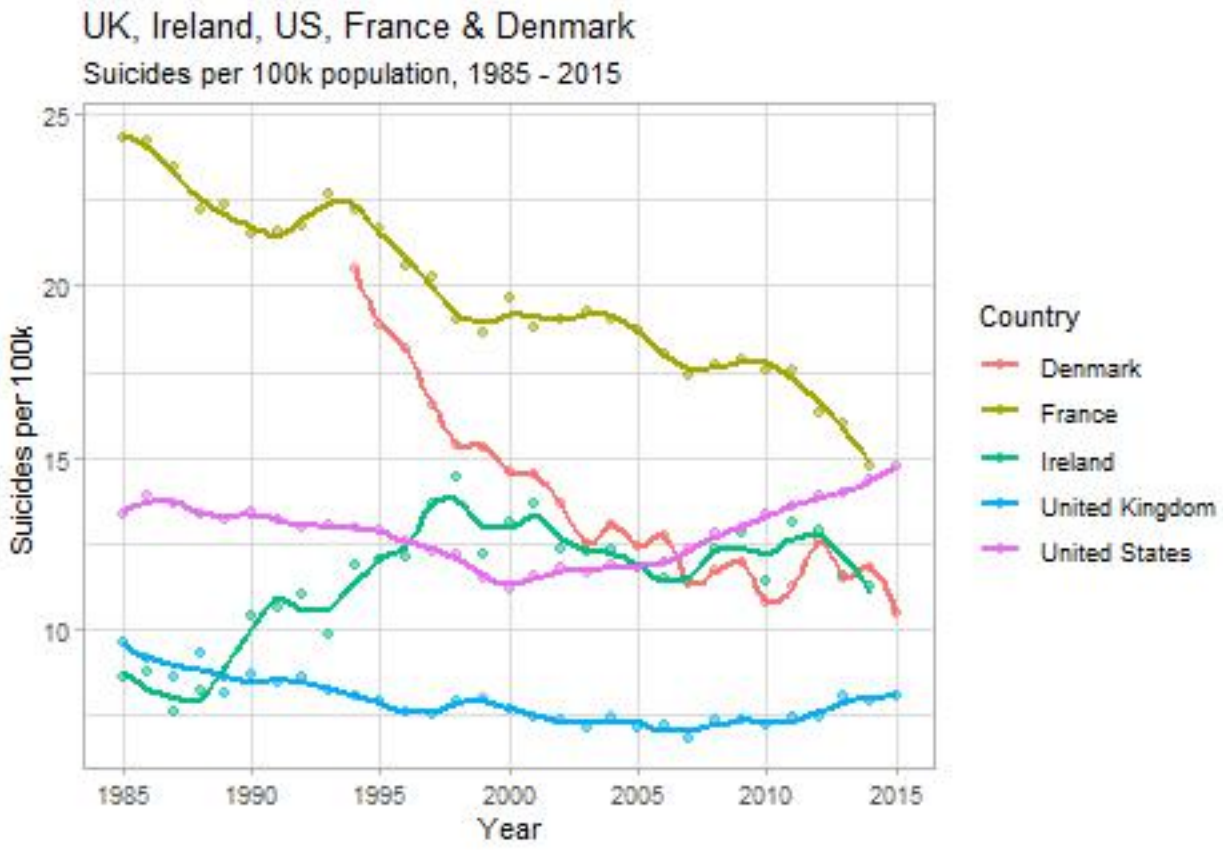
This was highlighted by my statistical analysis in **(2.5.2)**, which identified South Korea as the steepest increasing country, and Hungary as the 4th steepest decreasing country overall.

Comparing the UK, Ireland, America, France & Denmark

I think it would be useful to compare a few countries that people might think of as similar to the UK (culturally, legally, economically).

Overall Trend

```
data_filtered <- data %>%
  filter(country %in% c("United Kingdom",
                        "Ireland",
                        "United States",
                        "France",
                        "Denmark"))


data_filtered %>%
  group_by(country, year) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5) %>%
  ggplot(aes(x = year, y = suicide_per_100k,
             col = country)) +
  geom_point(alpha = 0.5) +
  geom_smooth(se = F, span = 0.2) +
  scale_x_continuous(breaks = seq(1985, 2015, 5),
                     minor_breaks = F) +
  labs(title = "UK, Ireland, US, France & Denmark",
       subtitle = "Suicides per 100k population, 1985 - 2015",
       x = "Year",
       y = "Suicides per 100k",
       col = "Country")
```



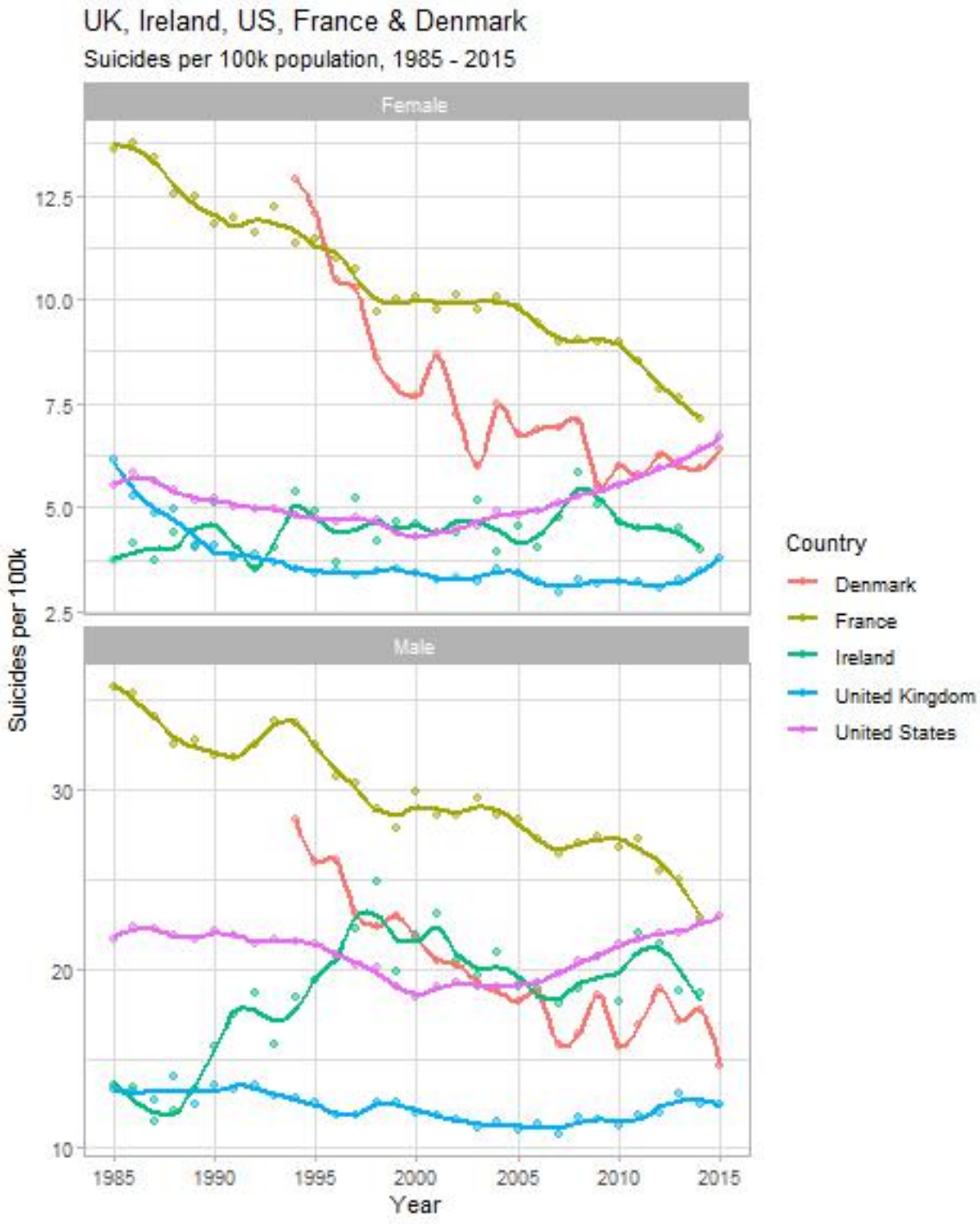
Insights

- The **UK suicide rate has been consistently lowest since 1990**, and has remained fairly static since ~1995
- France has historically had the highest rate, but is now roughly equal with America
- The **US has the most concerning trend, linearly increasing by ~1/3 since 2000**

By Sex

Male & Female Rates (over time)

```
data_filtered %>%
  group_by(country, sex, year) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5) %>%
  ggplot(aes(x = year, y = suicide_per_100k,
             col = country)) +
  geom_point(alpha = 0.5) +
  geom_smooth(se = F, span = 0.2) +
  scale_x_continuous(breaks = seq(1985, 2015, 5),
                     minor_breaks = F) +
  facet_wrap(~ sex, scales = "free_y", nrow = 2) +
  labs(title = "UK, Ireland, US, France & Denmark",
       subtitle = "Suicides per 100k population, 1985 - 2015",
       x = "Year",
       y = "Suicides per 100k",
       col = "Country")
```



Insights

- For the **UK**, there's **no obvious increase in the suicide rate for men than can't also be observed to an equal extent in women**
- Again, for men and women, **France has decreased** to being roughly equal with the US in 2015

- The different trend lines for men & women in **Ireland** is unusual - in **1990**, the **male rate increases**, but the **same can't be observed for females**

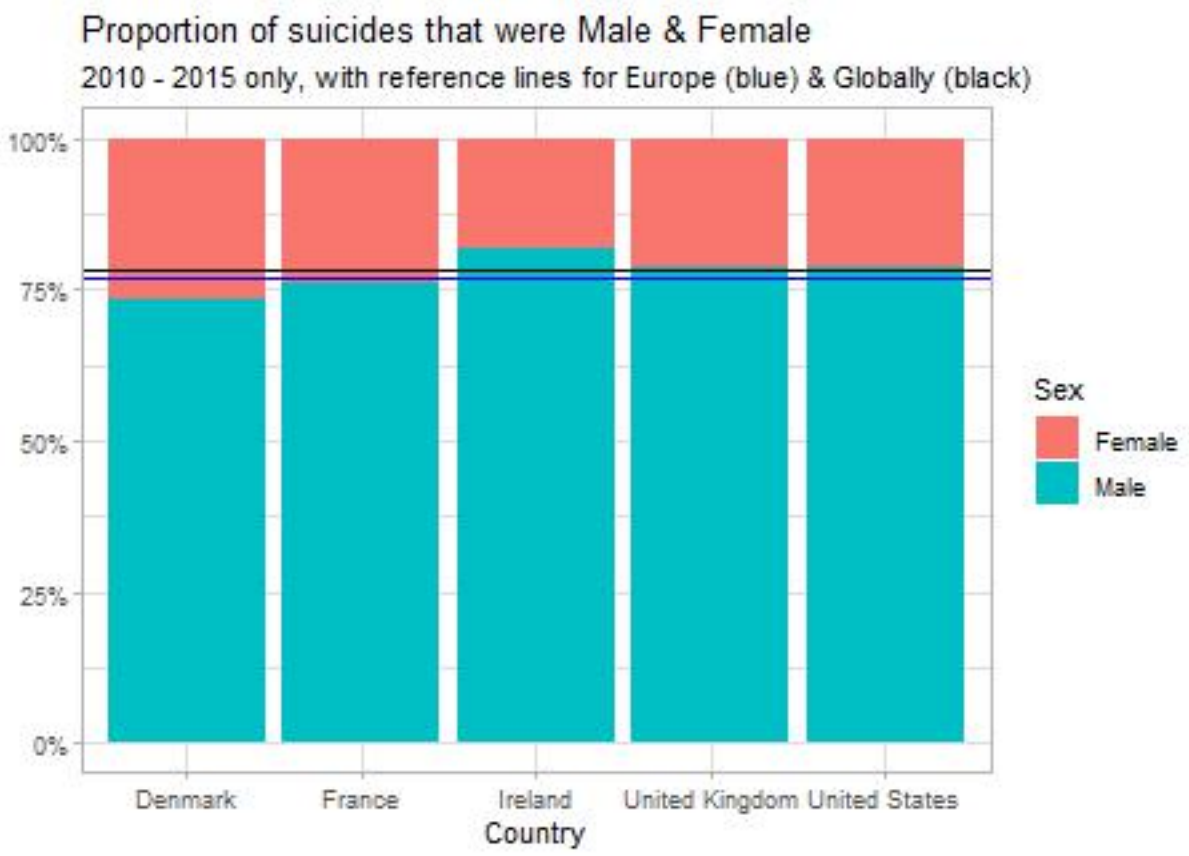
2010 - 2015 Only

For the purposes of these visualisations, i'm really more interested in data from recent years (France, for example, has changed a *lot*), so i'll **restrict the timeframe to 2010 onwards**.

Proportion of suicides that are Men

```
t1 <- data_filtered %>%
  filter(year >= 2010) %>%
  group_by(sex) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5)

global_male_proportion <- t1$suicide_per_100k[2] / sum(t1$suicide_per_100k)
#0.7802372


t2 <- data_filtered %>%
  filter(year >= 2010, continent == "Europe") %>%
  group_by(sex) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5)

european_male_proportion <- t2$suicide_per_100k[2] / sum(t2$suicide_per_100k)


data_filtered %>%
  filter(year >= 2010) %>%
  group_by(country, sex) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5) %>%
  ggplot(aes(x = country, y = suicide_per_100k,
             fill = sex)) +
  geom_bar(position = "fill", stat = "identity") +
  geom_hline(yintercept = global_male_proportion) +
  geom_hline(yintercept = european_male_proportion,
             col = "blue") +
  scale_y_continuous(labels = scales::percent) +
  labs(title = "Proportion of suicides that were Male & Female",
       subtitle = "2010 - 2015 only, with reference lines for Europe (blue) & Globally (black)",
       x = "Country",
       y = "",
       fill = "Sex")
```

```
## `summarise()` has grouped output by 'country'. You can override using the
## `.groups` argument.
```



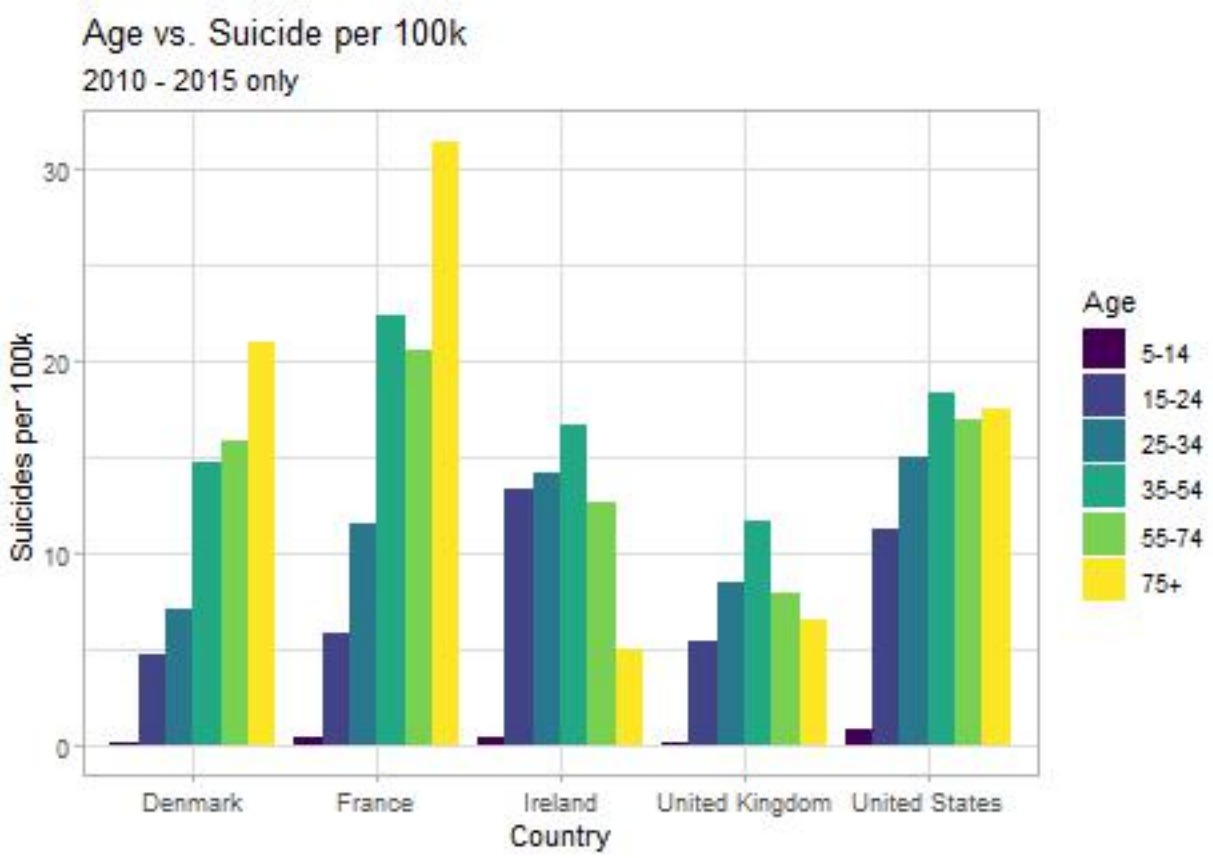
Insights

- Similar pattern as seen throughout the analysis - **men make up ~ 75% of deaths by suicide**
- The **highest** proportion is in **Ireland - 81.7% male**
- The **lowest** proportion is for **Denmark - 73.5% male**

Age Rates

```
data_filtered %>%
  filter(year >= 2010) %>%
  group_by(country, age) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5) %>%
  ggplot(aes(x = country, y = suicide_per_100k,
             fill = age)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title    = "Age vs. Suicide per 100k",
       subtitle = "2010 - 2015 only",
       x        = "Country",
       y        = "Suicides per 100k",
       fill     = "Age")
```

```
## `summarise()` has grouped output by 'country'. You can override using the
## `.groups` argument.
```



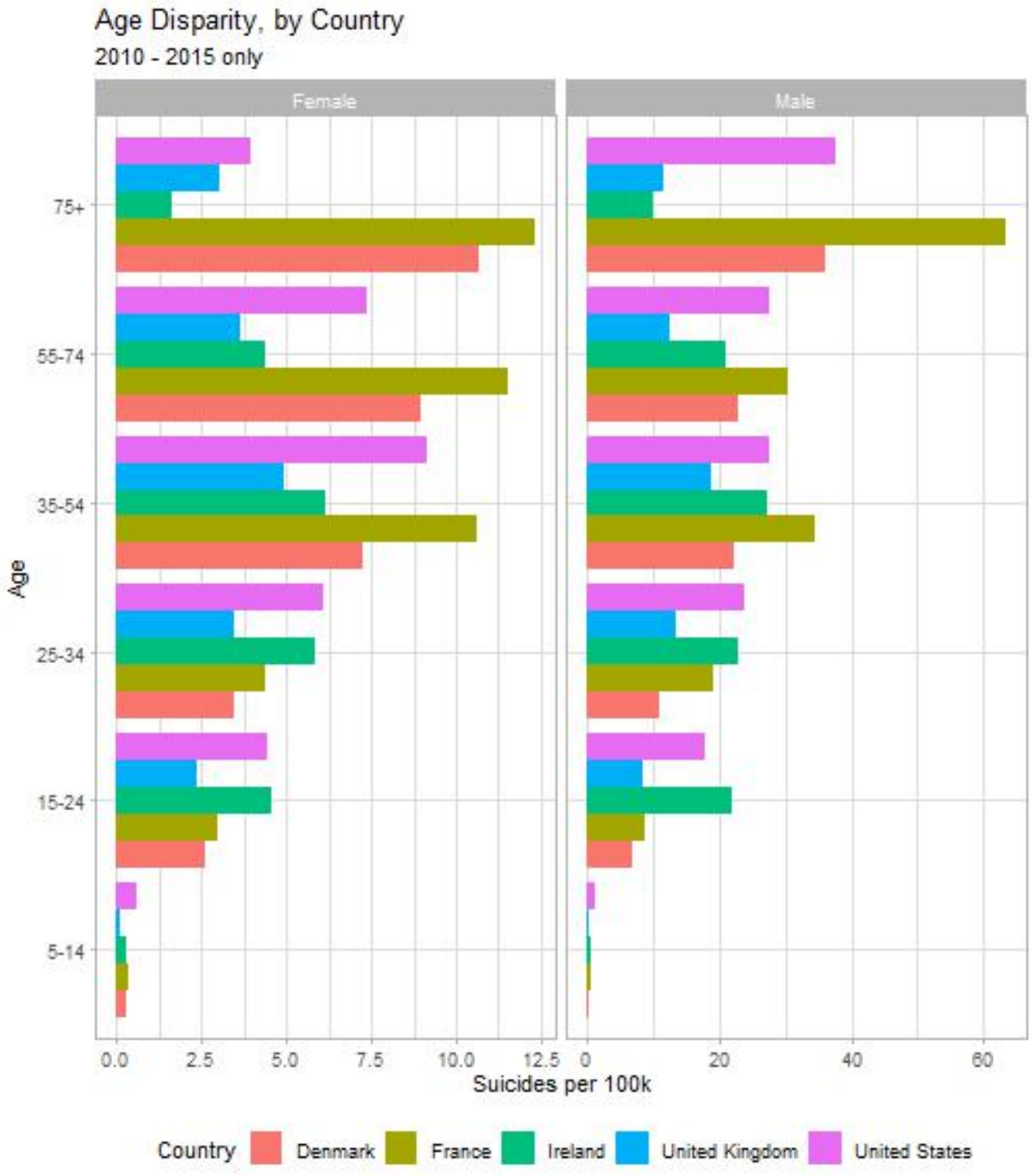
Insights

- There's a huge difference in the 'trend' of suicide rates as age varies within each country
- **Suicide rate increases with age for France, Denmark and the US** (to a lesser extent)
- Those aged **35-54** at the **highest risk** in **Ireland and the UK**, which follow closer to a gaussian distribution

Male & Female Rates (for different age categories)

```
data_filtered %>%
  filter(year >= 2010) %>%
  group_by(country, sex, age) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5) %>%
  ggplot(aes(x = age, y = suicide_per_100k,
             fill = country)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_wrap(~ sex, scales = "free_x") +
  labs(title = "Age Disparity, by Country",
       subtitle = "2010 - 2015 only",
       x        = "Age",
       y        = "Suicides per 100k",
       fill = "Country") +
  coord_flip() +
  theme(legend.position = "bottom")
```

```
## `summarise()` has grouped output by 'country', 'sex'. You can override using
## the `.groups` argument.
```



Insights

- In the **US**, **suicide rate for men continues to increase with age**, but the **female rate decreases in old age**
- This weird disparity is only present in the US and i'm curious as to why it occurs
- The **UK has the lowest or second lowest suicide rate in every sex-age group**

Young to Middle-Aged Men

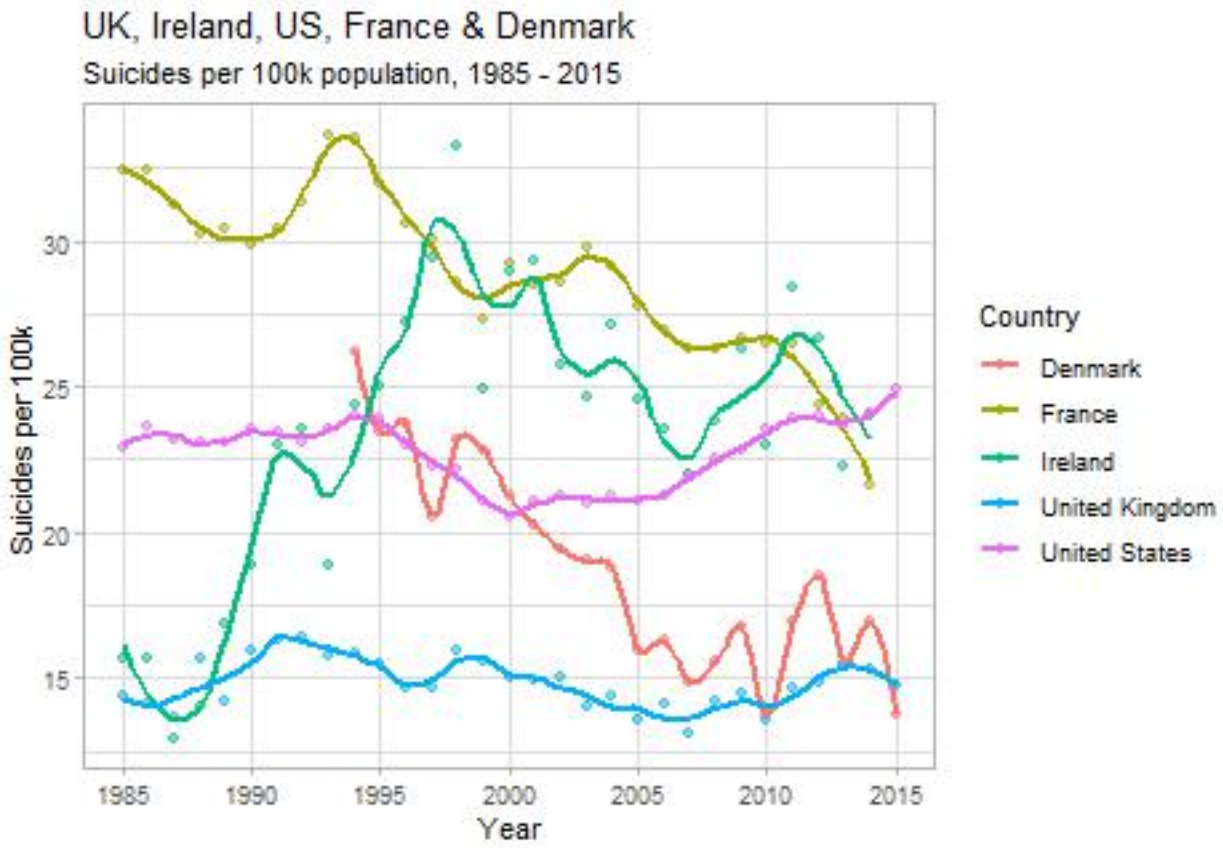
There is a big concern in my country (UK) regarding mental health problems and suicide for young to middle-aged men. Here i'm going to **restrict the analysis** to just:

- **Men**
- **Ages "15-24", "25-34" & "35-54"**

I'll basically be observing whether concerning trends are present. I think having other countries here for comparison will be useful and will help provide perspective in the analysis.

Men - Ages 15-54 Combined

```
data_filtered %>%
  filter(age %in% c("15-24", "25-34", "35-54"), sex == "Male") %>%
  group_by(country, year) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) /
                                  sum(as.numeric(population))) * 10^5) %>%
  ggplot(aes(x = year, y = suicide_per_100k,
             col = country)) +
  geom_point(alpha = 0.5) +
  geom_smooth(se = F, span = 0.2) +
  scale_x_continuous(breaks = seq(1985, 2015, 5),
                     minor_breaks = F) +
  labs(title = "UK, Ireland, US, France & Denmark",
       subtitle = "Suicides per 100k population, 1985 - 2015",
       x = "Year",
       y = "Suicides per 100k",
       col = "Country")
```

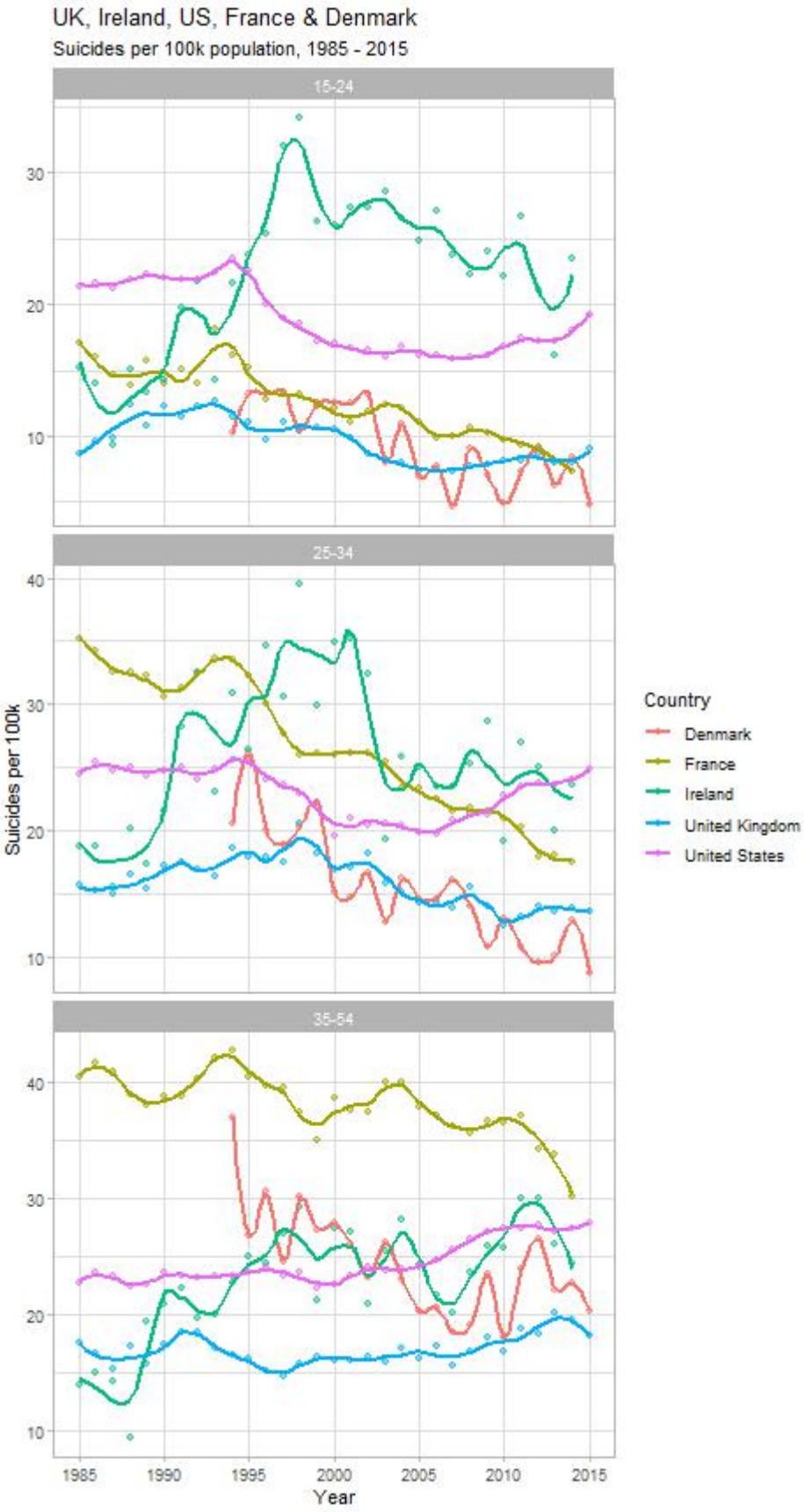


Insights

- **Ireland's trend over the 1990's was very concerning**
- It went from 14 (per 100k, per year) to 33.3 between 1988 and 1998 - an **increase of 138%**
- Again, the **US** shows the most obvious and **concerning current trend**
- *Comparatively*, for young to middle-aged men, the **UK seems fairly flat across time**

Men - Ages 15-24, 25-34 & 35-54

```
data_filtered %>%
  filter(age %in% c("15-24", "25-34", "35-54"), sex == "Male") %>%
  group_by(country, age, year) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) / sum(as.numeric(population))) * 100000) %
  ggplot(aes(x = year, y = suicide_per_100k,
             col = country)) +
  geom_point(alpha = 0.5) +
  geom_smooth(se = F, span = 0.2) +
  facet_wrap(~ age, nrow = 3,
             scales = "free_y") +
  scale_x_continuous(breaks = seq(1985, 2015, 5),
                     minor_breaks = F) +
  labs(title = "UK, Ireland, US, France & Denmark",
       subtitle = "Suicides per 100k population, 1985 - 2015",
       x = "Year",
       y = "Suicides per 100k",
       col = "Country")
```


UK, Ireland, US, France & Denmark
Suicides per 100k population, 1985 - 2015
15-24
25-34
35-54
Suicides per 100k
Year
1985
1990
1995
2000
2005
2010
2015
10
20
30
40
Country
Denmark
France
Ireland
United Kingdom
United States

Insights

- For **men in the UK**, only the **'35-54'** category seems to be increasing, with a **slight increase (~10-15%)** over the past decade
- **UK** rates for men in the **'15-24'** and **'25-34'** categories appear **flat** & **slightly decreasing**, respectively
- It has been quite hard to describe these UK trends, which I think is a **positive** thing. The comparison with other countries definitely helps with perspective

Conclusion

- **Global Trends**: The global suicide rate decreased by about 25% from its peak in 1995 to 2015.
- **Continental Variations** : Europe initially had the highest suicide rate but experienced a significant decline. Oceania and the Americas showed more concerning trends, with increasing or stagnant rates.
- **Gender and Age**: The male suicide rate was consistently 3.5 times higher than the female rate. The likelihood of suicide generally increased with age, with a significant drop observed in the 75+ age group since 1990.
- **Country-Specific Analysis**: Lithuania had the highest suicide rate, while the UK consistently had the lowest. The analysis also highlighted specific trends in countries like South Korea (increasing rates) and Hungary (decreasing rates).
- **Gender Disparity**: The over-representation of men in suicide deaths was observed globally and across all continents and countries. The gender paradox in suicidal behavior was noted, with men being more likely to die from suicide despite women having higher rates of depression and suicidal thoughts.
- **Age Disparity by Continent**: The analysis revealed that suicide rates generally increase with age in the Americas, Asia, and Europe. However, in Oceania and Africa, the highest rates were found in the 25-34 age group.
- **Relationship between GDP and Suicide Rates**: The analysis found a weak but significant positive relationship between a country's GDP per capita and its suicide rate.
- **The 5% Highest Risk Instances**: The analysis identified demographics with the highest suicide rates, highlighting the over-representation of males and the 75+ age group in these instances. South Korea and Hungary were identified as having consistently high-risk demographics for men.
- **Comparison of UK, Ireland, US, France, and Denmark**: The UK had the lowest suicide rate among these countries, while the US showed the most concerning trend with increasing rates. The analysis also explored gender and age-specific trends within these countries.

Thank you :D