

Covid 19 plots

import data

```
library(tidyverse)
```

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

```
## v ggplot2 3.3.5      v purrr 0.3.4
## v tibble 3.1.2       v dplyr 1.0.6
## v tidyr 1.1.4        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()
```

```
covid <- read.csv("covid_19_data.csv")
str(covid)
```

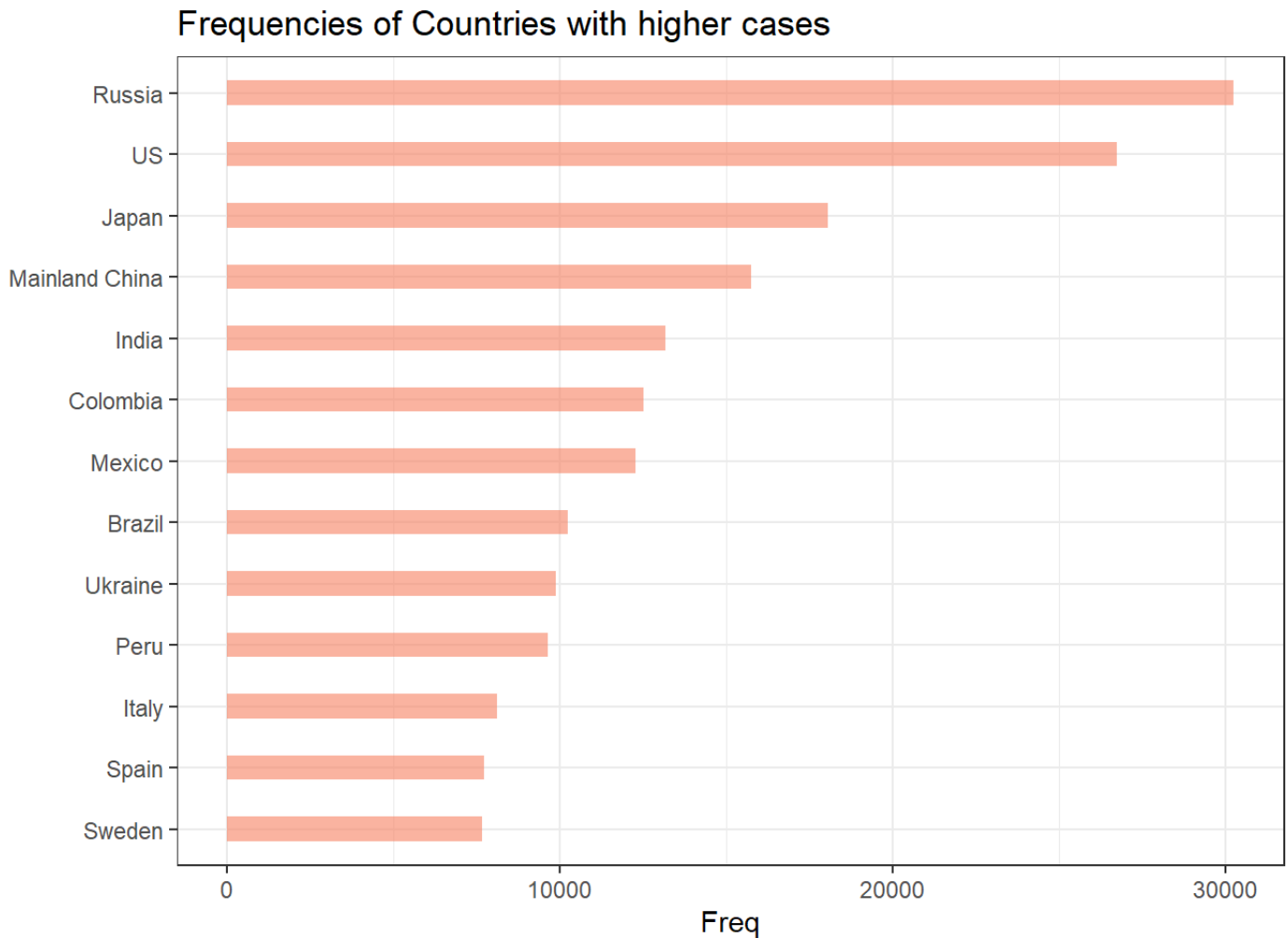
```
## 'data.frame': 306429 obs. of 8 variables:
## $ SNo : int 1 2 3 4 5 6 7 8 9 10 ...
## $ ObservationDate: chr "01/22/2020" "01/22/2020" "01/22/2020" "01/22/2020" ...
## $ Province.State: chr "Anhui" "Beijing" "Chongqing" "Fujian" ...
## $ Country.Region: chr "Mainland China" "Mainland China" "Mainland China" "Mainl
and China" ...
## $ Last.Update : chr "1/22/2020 17:00" "1/22/2020 17:00" "1/22/2020 17:00" "1/
22/2020 17:00" ...
## $ Confirmed : num 1 14 6 1 0 26 2 1 4 1 ...
## $ Deaths : num 0 0 0 0 0 0 0 0 0 0 ...
## $ Recovered : num 0 0 0 0 0 0 0 0 0 0 ...
```

barplot of country region

```

country_tab <- table(covid$Country.Region)
country_tab <- data.frame(country_tab)
country_tab %>% mutate(region = fct_reorder(Var1, Freq)) %>%
  filter(Freq>7000) %>%
  ggplot( aes(x=region, y=Freq)) +
  geom_bar(stat="identity", fill="#f68060", alpha=.6, width=.4) +
  ggtitle("Frequencies of Countries with higher cases")+
  coord_flip() +
  xlab("") +
  theme_bw()

```



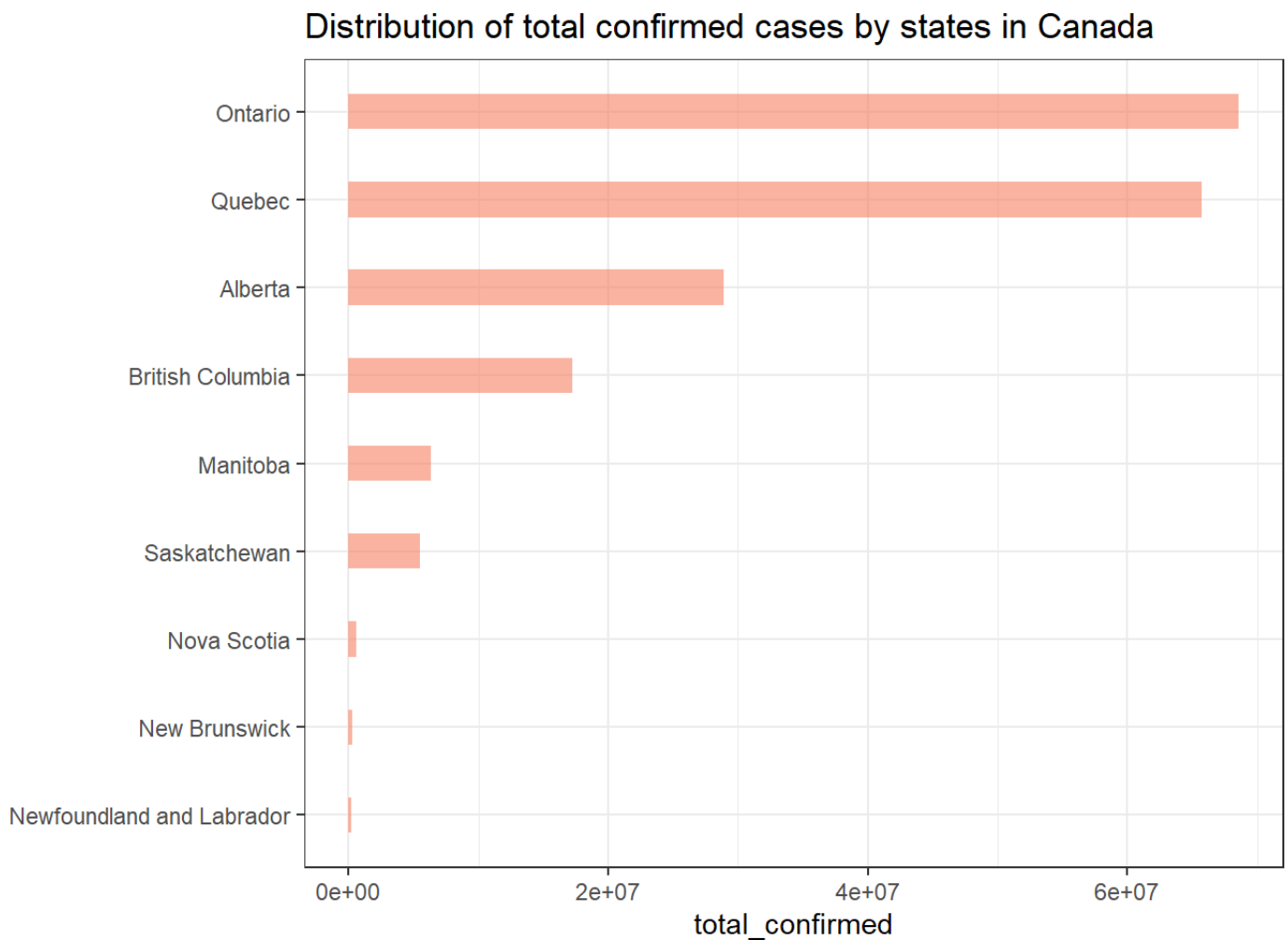
The bar graph shows the frequencies of country regions and it is evident that Russia had most of the regions where covid cases were.

```

can_dat <- covid %>% filter(Country.Region == "Canada")

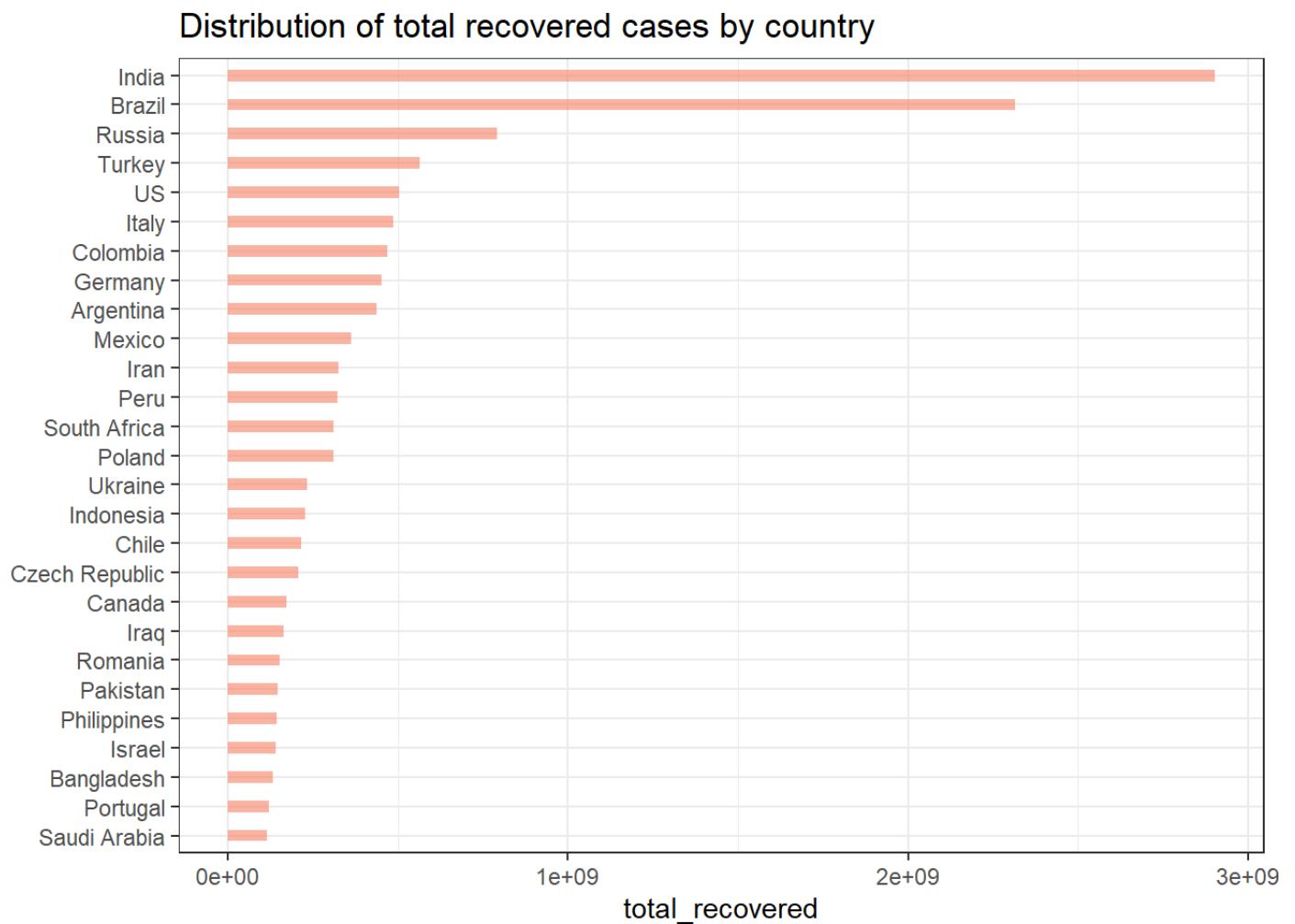
can_dat %>% group_by(Province.State) %>%
  summarise(total_confirmed = sum(Confirmed)) %>%
  mutate(province = fct_reorder(Province.State, total_confirmed)) %>%
  filter(total_confirmed>100000) %>%
  ggplot(aes(province, total_confirmed))+
  geom_bar(stat="identity", fill="#f68060", alpha=.6, width=.4) +
  ggtitle("Distribution of total confirmed cases by states in Canada")+
  coord_flip() +
  xlab("") +
  theme_bw()

```



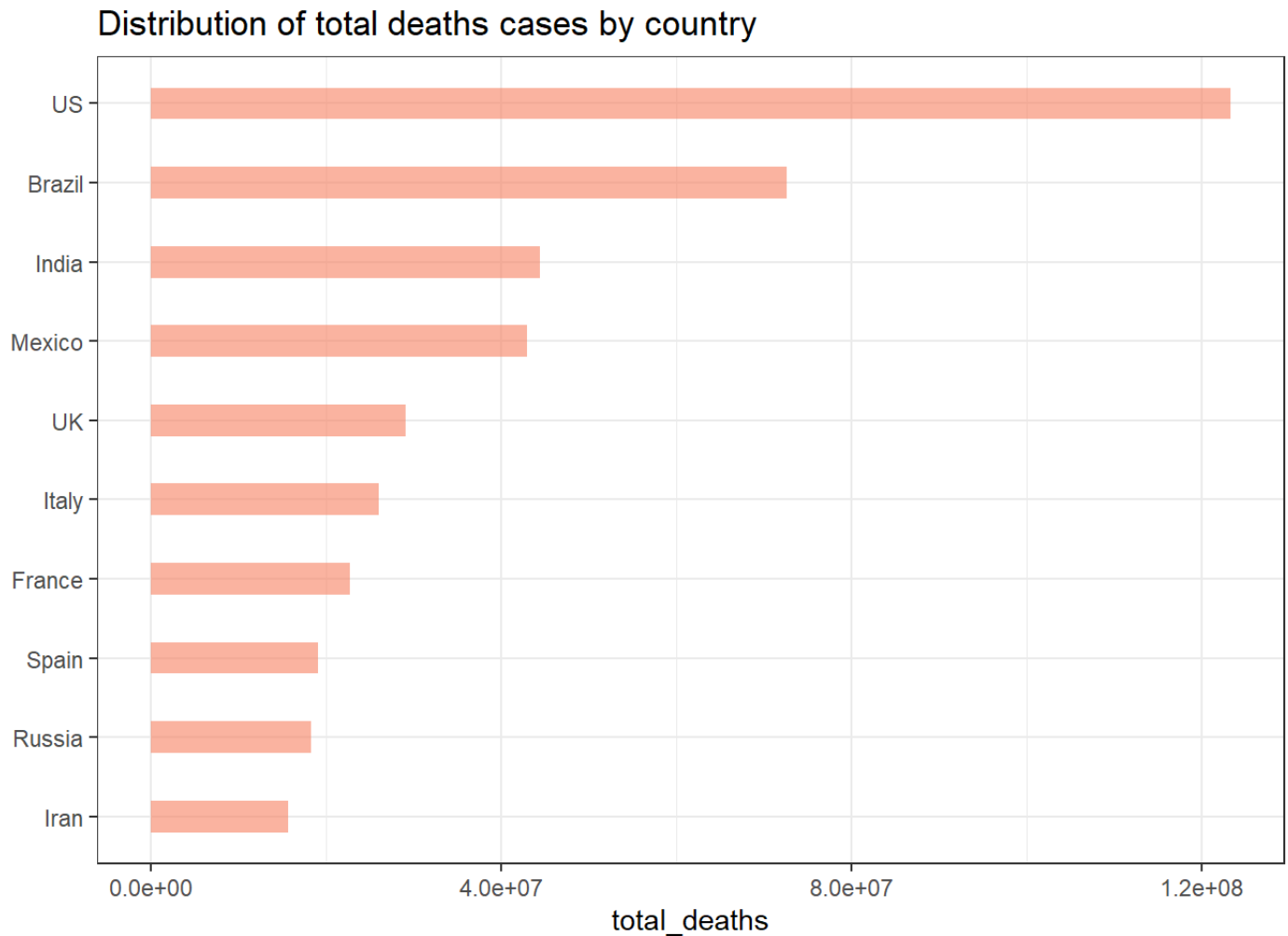
The bar graph indicates that Ontario was the province with higher cases then followed by Quebec and then Alberta in Canada.

```
covid %>% group_by(Country.Region) %>%
  summarise(total_recovered = sum(Recovered)) %>%
  mutate(country = fct_reorder(Country.Region, total_recovered)) %>%
  filter(total_recovered>100000000) %>%
  ggplot(aes(country, total_recovered))+
  geom_bar(stat="identity", fill="#f68060", alpha=.6, width=.4) +
  ggtitle("Distribution of total recovered cases by country")+
  coord_flip() +
  xlab("") +
  theme_bw()
```



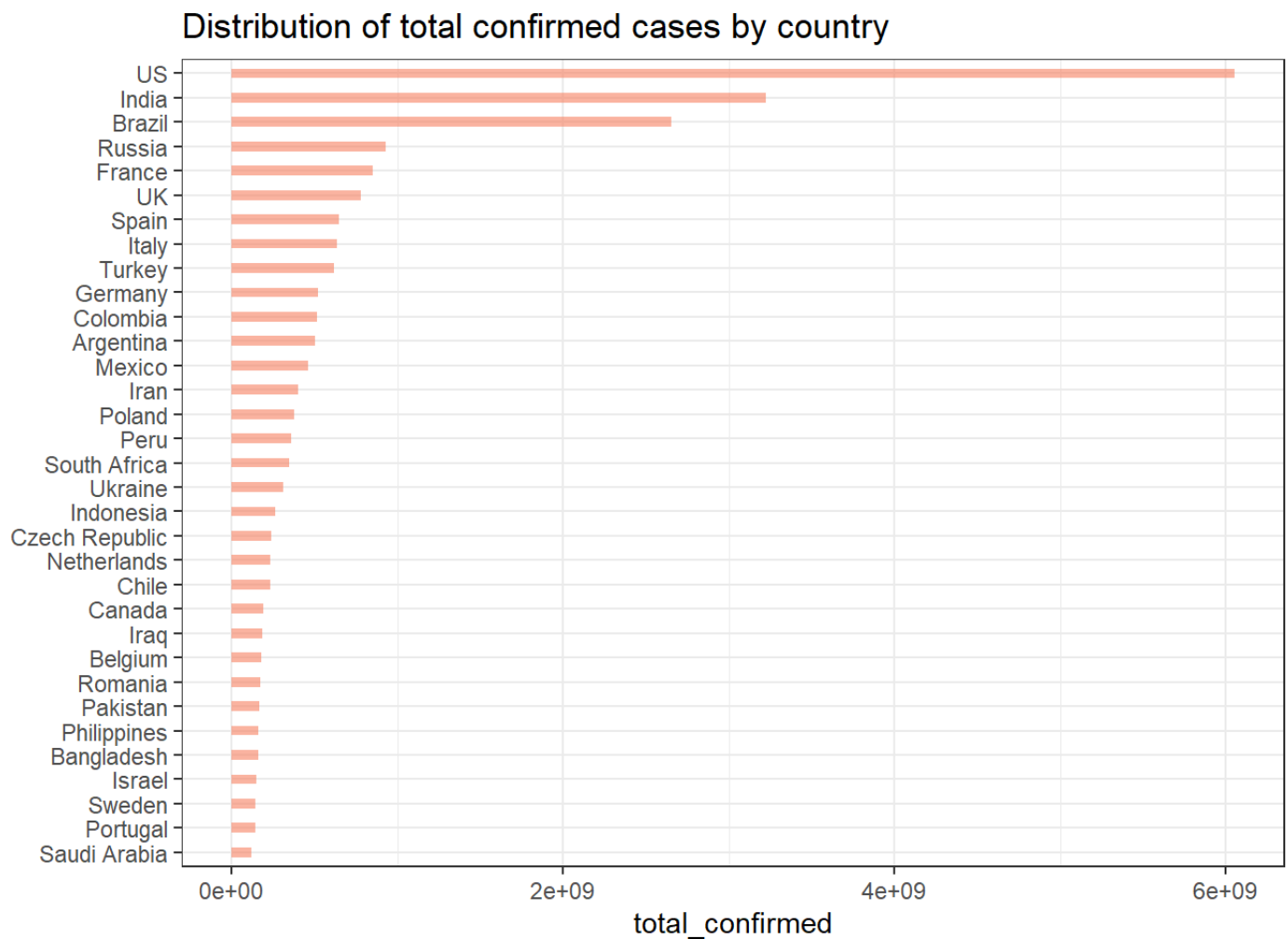
The bar graph gives the total recovered cases per country, it is evident that India had the highest recovery cases then followed by Brazil and Russia.

```
covid %>% group_by(Country.Region) %>%
  summarise(total_deaths = sum(Deaths)) %>%
  mutate(country = fct_reorder(Country.Region, total_deaths)) %>%
  filter(total_deaths>15000000) %>%
  ggplot(aes(country, total_deaths))+
  geom_bar(stat="identity", fill="#f68060", alpha=.6, width=.4) +
  ggtitle("Distribution of total deaths cases by country")+
  coord_flip() +
  xlab("") +
  theme_bw()
```



By that time of the year US had the highest total deaths then followed by Brazil and India.

```
covid %>% group_by(Country.Region) %>%
  summarise(total_confirmed = sum(Confirmed)) %>%
  mutate(country = fct_reorder(Country.Region, total_confirmed)) %>%
  filter(total_confirmed>120000000) %>%
  ggplot(aes(country, total_confirmed))+
  geom_bar(stat="identity", fill="#f68060", alpha=.6, width=.4) +
  ggtitle("Distribution of total confirmed cases by country")+
  coord_flip() +
  xlab("") +
  theme_bw()
```



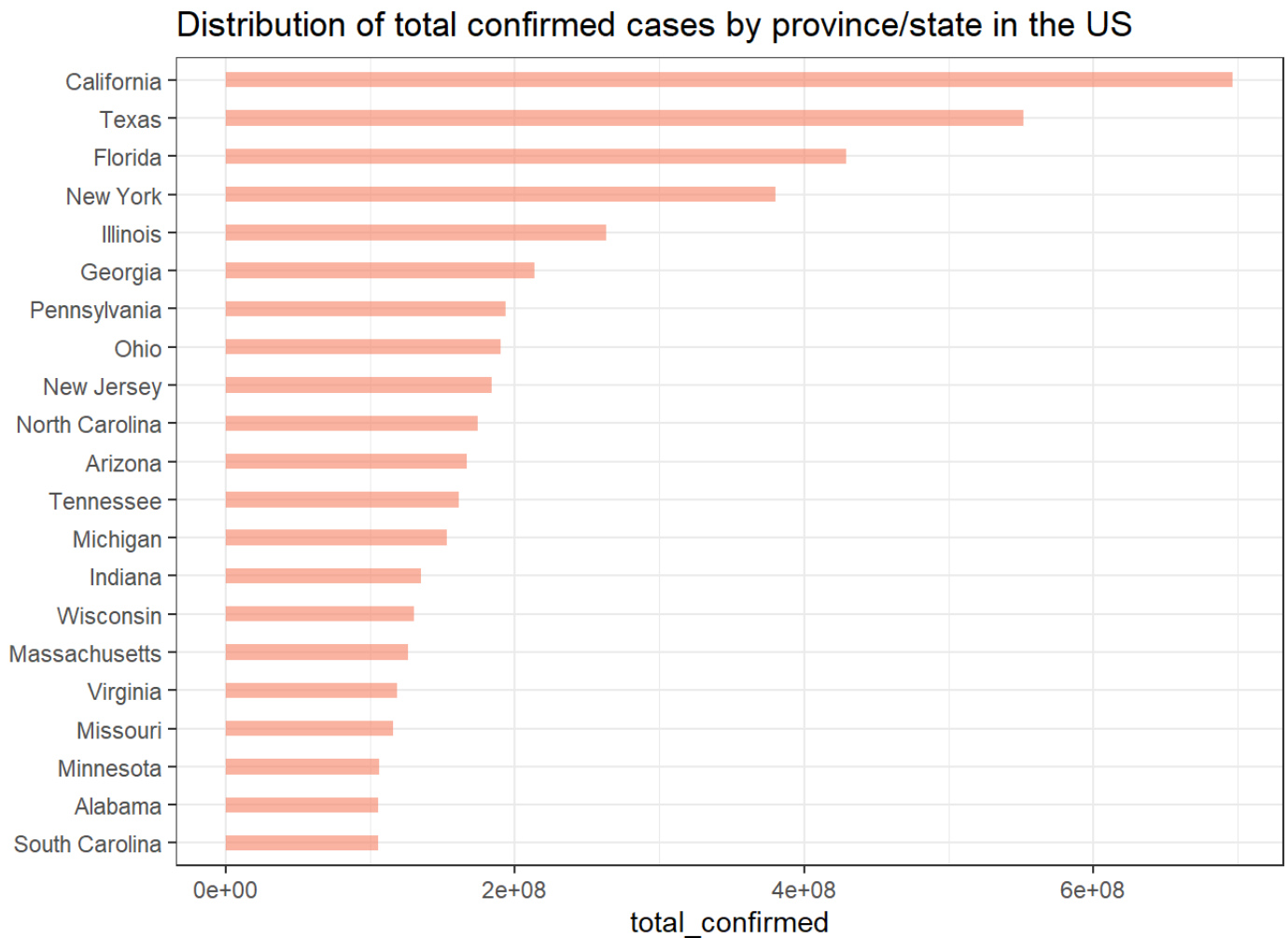
The Bar graph shows the total number of confirmed cases as at that time of the year with United States having the highest number of confirmed cases the followed by India and Brazil.

```

US_dat <- covid %>% filter(Country.Region == "US")

US_dat %>% group_by(Province.State) %>%
  summarise(total_confirmed = sum(Confirmed)) %>%
  mutate(province = fct_reorder(Province.State, total_confirmed)) %>%
  filter(total_confirmed>100000000) %>%
  ggplot(aes(province, total_confirmed))+
  geom_bar(stat="identity", fill="#f68060", alpha=.6, width=.4) +
  ggtitle("Distribution of total confirmed cases by province/state in the US")+
  coord_flip() +
  xlab("") +
  theme_bw()

```



Narrowing the study and analysis to one country leads us to The United states and we analyse the total confirmed cases by states and it shows that California had the most confirmed cases followed by Texas and then Florida.