Scraping data from google_cmr

yakov

20/11/2020

Task 1

```
glob_mob1=read.csv("https://www.gstatic.com/covid19/mobility/Global_Mobility_Report.csv")
```

```
glob_mob=glob_mob1
```

Here I subset out the "Whole country" parts of the data. So no subregion data, only data for the whole country.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
glob_mob=magrittr::extract(glob_mob,!names(glob_mob) %in% c("census_fips_code","sub_region_1","sub_regi
```

Now I do the function from the teacher, which he gave us on the labs:

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
glob_mob$date = as.Date(parse_date_time(glob_mob$date,orders=c("y","ym","ymd")))
```

```
glob_mob<- {glob_mob %>%
    filter(date > as.Date("2020-02-06"))}
```

```
europe_list=c("Albania","Andorra","Armenia","Austria","Azerbaijan","Belarus","Belgium","Bosnia & Herzeg

NA_list=c("United States","Mexico","Canada","Guatemala","Cuba","Haiti","Dominican Republic","Honduras",

SA_list=c("Aruba","Brazil","Colombia","Argentina","Peru","Venezuela","Chile","Ecuador","Bolivia","Parag
```

```
glob_mob= glob_mob%>%
  mutate(country_region_code=ifelse(country_region %in% NA_list,"North America",ifelse(country_region %

colnames(glob_mob)[colnames(glob_mob) == "country_region_code"] <- "region"

colnames(glob_mob)[colnames(glob_mob) == "country_region"] <- "country"



glob_mob=glob_mob[!(glob_mob$region=="NON"),]
```

Here I add the feature income from the dataset from the labs to the glob_mob dataset. I think now, the dataset glob_mob is ready to use

```
mean_ <- function(...) mean(..., na.rm=T)
max_ <- function(...) max(..., na.rm=T)
groceryData = {glob_mob %>%
    group_by(country) %>%
    summarise(mean_grocery_pharmacy = mean_(grocery_and_pharmacy_percent_change_from_baseline),
              mean_retail=mean_(retail_and_recreation_percent_change_from_baseline),
              region=region)}
```

```
## `summarise()` regrouping output by 'country' (override with `.groups` argument)
```

```
groceryData=unique(groceryData)
```

```
library(ggplot2)
glob_mob_small=glob_mob[1:nrow(glob_mob)/4,]
```

```
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

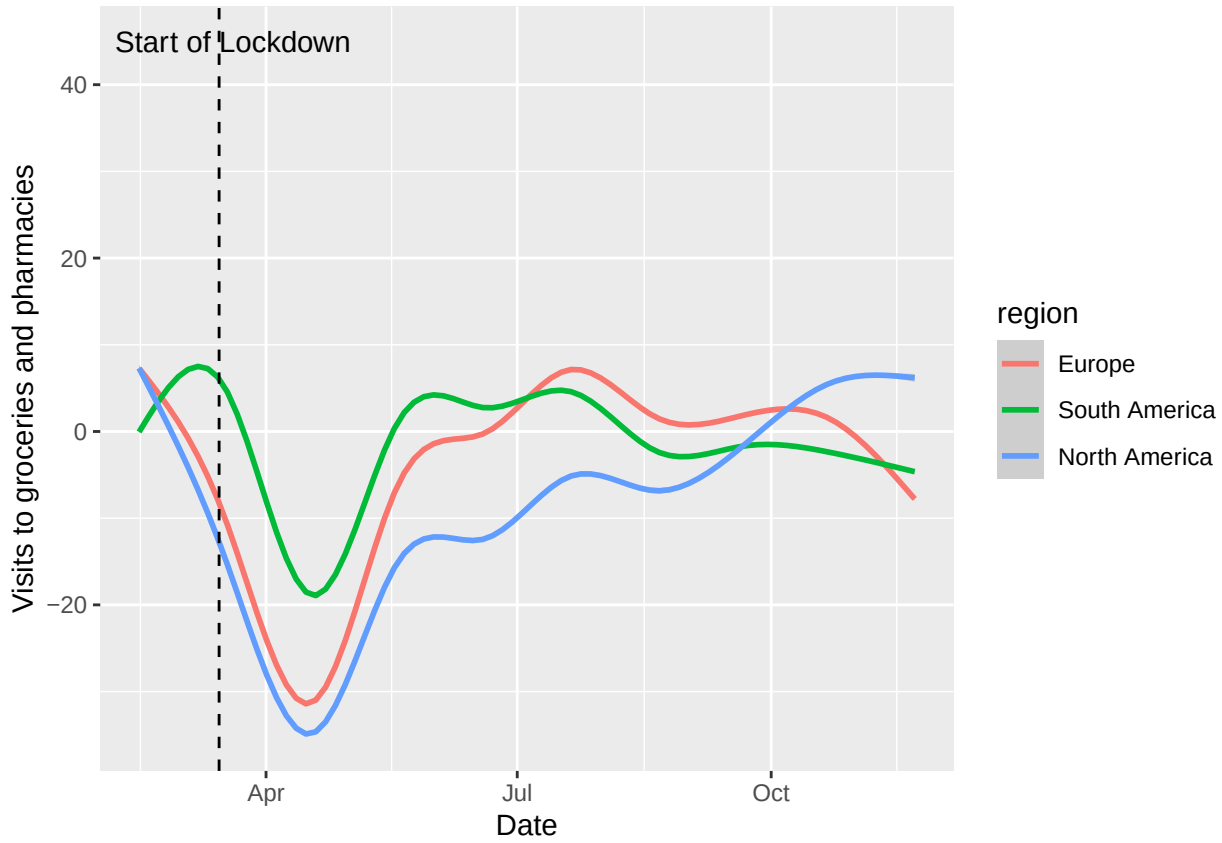
```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
plot1 <- {glob_mob %>% ggplot(aes(x=date,
                                  y=grocery_and_pharmacy_percent_change_from_baseline,color=r

print(plot1)
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

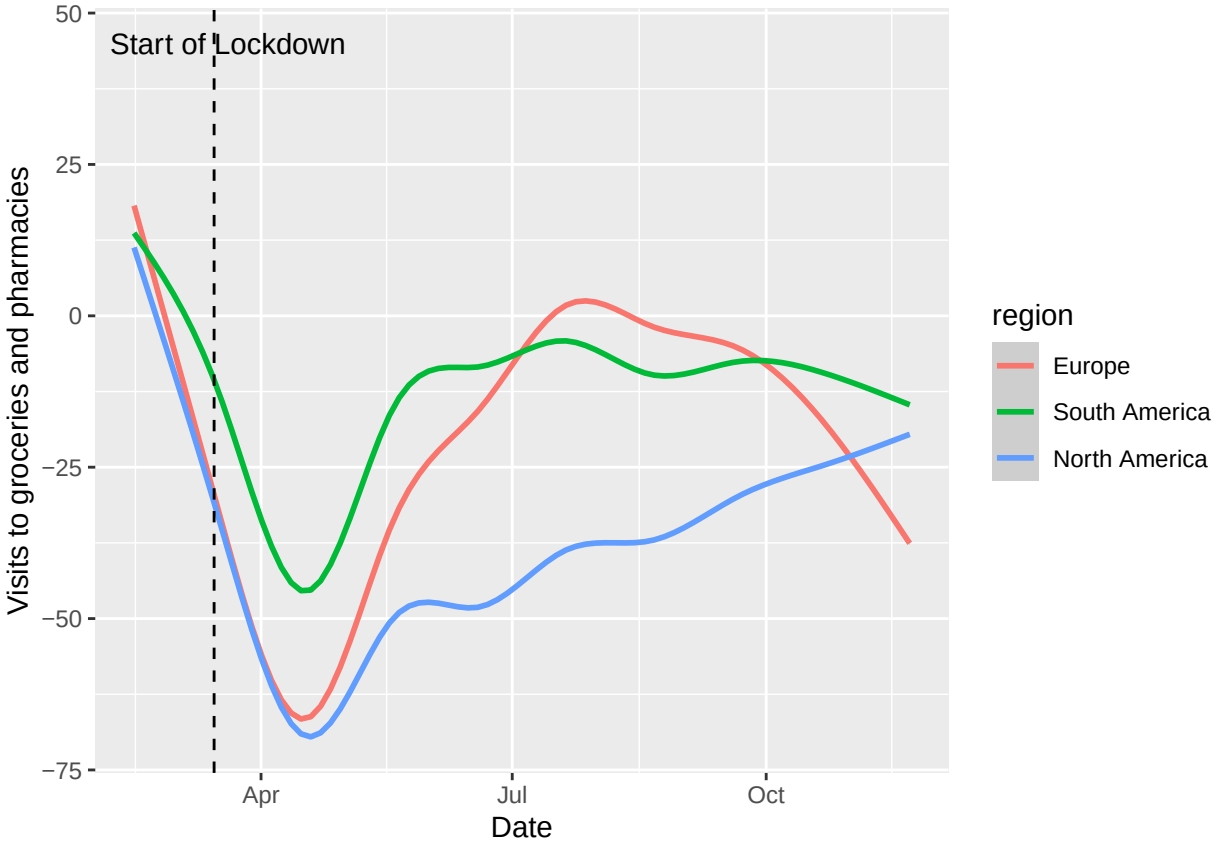
```
## Warning: Removed 1103644 rows containing non-finite values (stat_smooth).
```



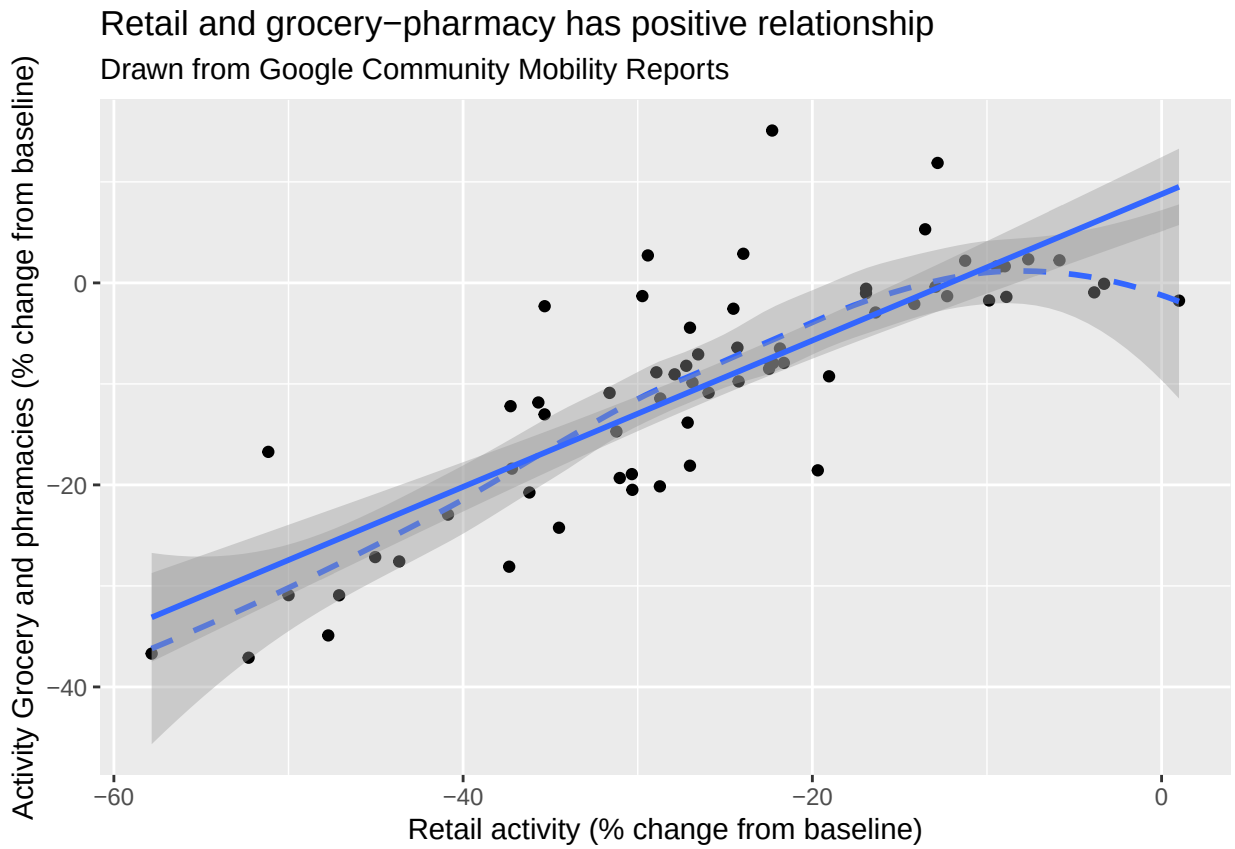

```
plot1 <- {glob_mob %>% ggplot(aes(x=date,
                                  y=retail_and_recreation_percent_change_from_baseline,color=

print(plot1)
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## Warning: Removed 1061437 rows containing non-finite values (stat_smooth).
```



We plot the average percentage change in the frequency of visits to grocery shops and pharmacies against retail and recreation for countries in Europe and America from this year (Feb 7 onwards). We look particularly at the trends for Europe, North America and South America. We fit a linear model as well as the best line of fit and see that a linear relationship between the variables fits fairly well.

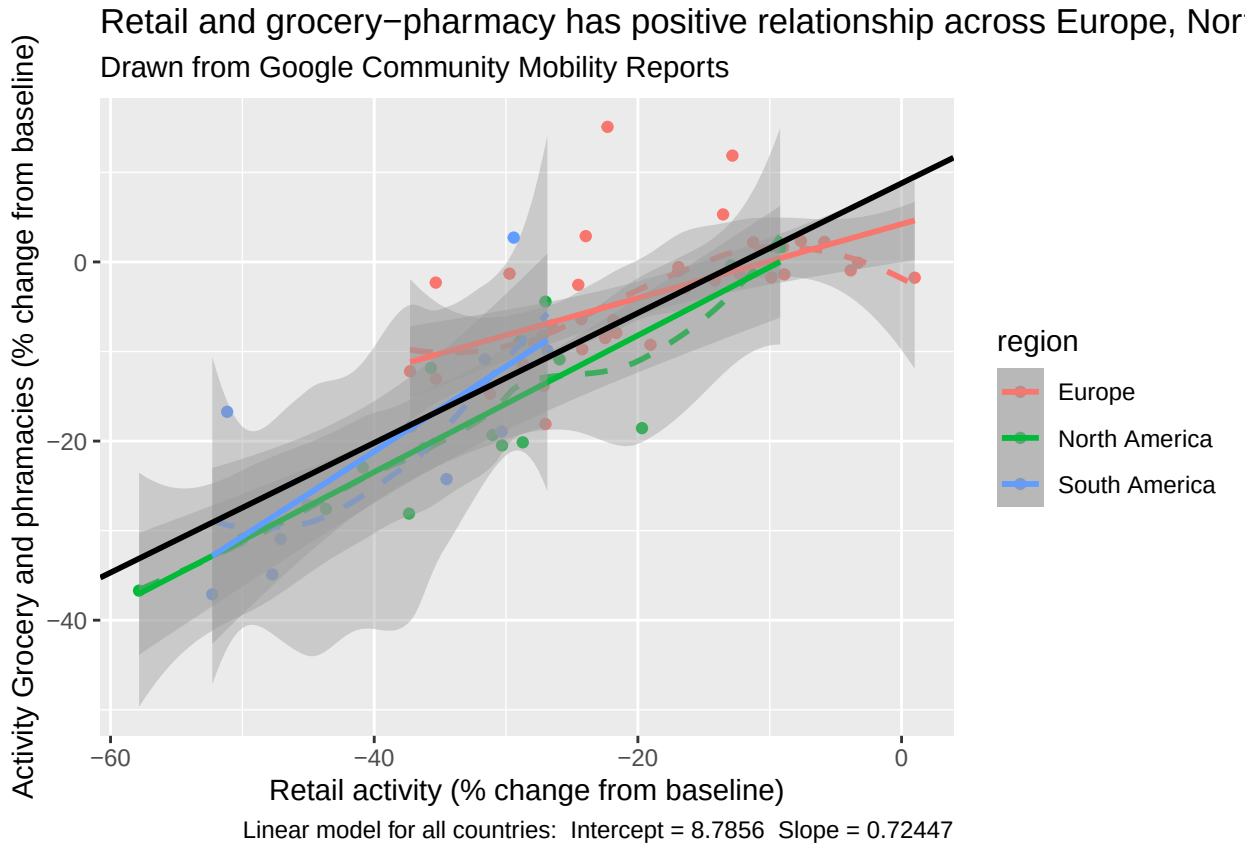



```
library(ggplot2)
lm.homework <- lm(mean_grocery_pharmacy ~ mean_retail, data = groceryData)


gro_ret_reg <- {groceryData %>% ggplot(aes(x = mean_retail,
                                           y = mean_grocery_pharmacy,
                                           color=region))} +
  xlab("Retail activity (% change from baseline)") +
  ylab("Activity Grocery and phramacies (% change from baseline)") +
  ggtitle("Retail and grocery-pharmacy has positive relationship across Europe, North America and South
          subtitle = "Drawn from Google Community Mobility Reports") +
  geom_point() +
  geom_smooth(method = loess, linetype = "dashed") +
  stat_smooth(method = "lm") +
  geom_abline(intercept = signif(lm.homework$coef[[1]],5),
              slope = signif(lm.homework$coef[[2]],5), color="black", size=1) +
  labs(caption = paste("Linear model for all countries: ",
                       "Intercept =",signif(lm.homework$coef[[1]],5),
                       " Slope =",signif(lm.homework$coef[[2]],5)))

print(gro_ret_reg)
```

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

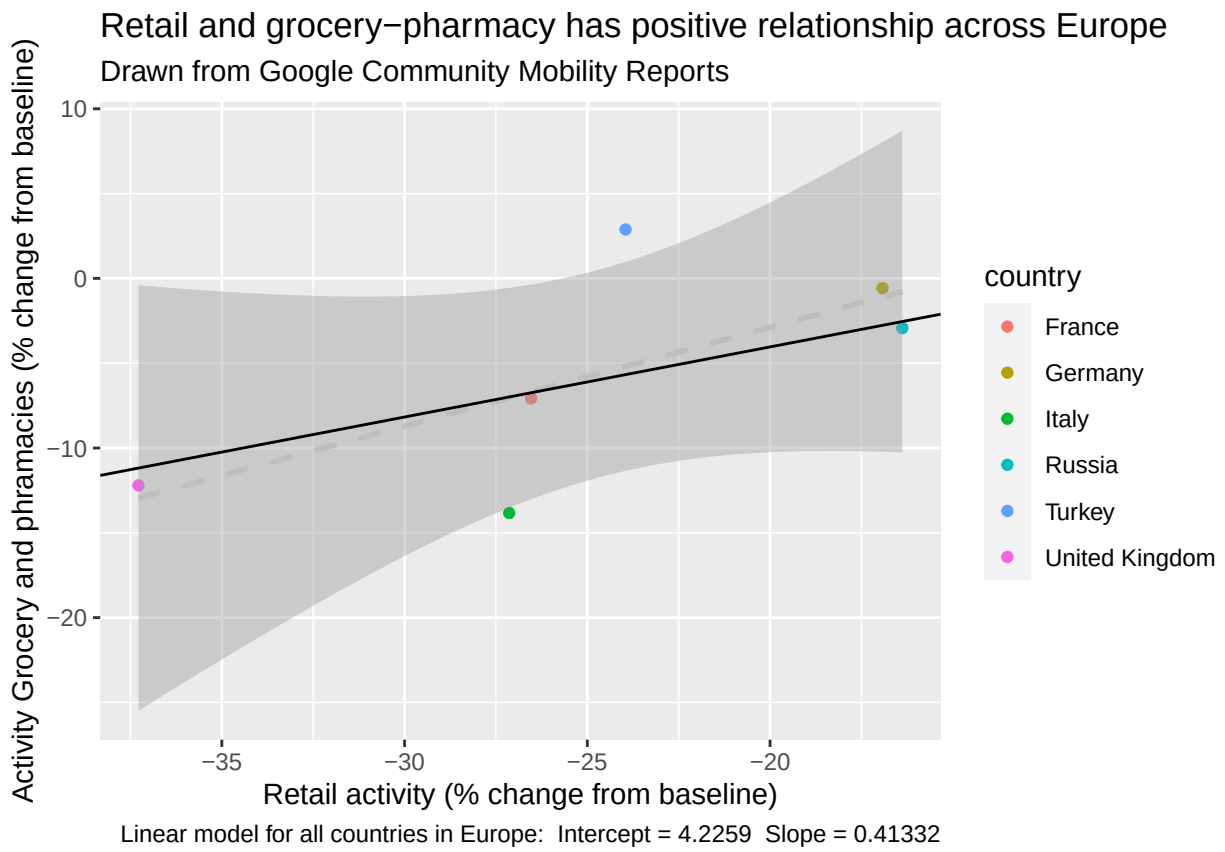



```
EUR <- {groceryData %>%
    filter(region == "Europe")}
N.USA <- {groceryData %>%
    filter(region == "North America")}
S.USA <- {groceryData %>%
    filter(region == "South America")}
EUR1 <- {EUR %>%
    filter(country %in% c("France","Germany","Italy","Russia","Turkey","United Kingdom"))}
N.USA1 <- {N.USA %>%
    filter(country %in% c("Canada","Cuba","Guatemala","Haiti","Mexico","United States"))}
S.USA1 <- {S.USA %>%
    filter(country %in% c("Argentina","Brazil","Chile","Colombia","Peru","Venezuela"))}
```

```
lm.EURhomework <- lm(mean_grocery_pharmacy ~ mean_retail, data = EUR)
EUR_home_work <- {EUR1 %>% ggplot(aes(x = mean_retail,
                                      y = mean_grocery_pharmacy,
                                      color=country))} +
  xlab("Retail activity (% change from baseline)") +
  ylab("Activity Grocery and phramacies (% change from baseline)") +
  ggtitle("Retail and grocery-pharmacy has positive relationship across Europe",
          subtitle = "Drawn from Google Community Mobility Reports") +
  geom_point() +
  stat_smooth(method = "lm", color="grey", linetype="dashed") +
  geom_abline(intercept = signif(lm.EURhomework$coef[[1]],5),
              slope = signif(lm.EURhomework$coef[[2]],5), color="black") +
  labs(caption = paste("Linear model for all countries in Europe: ",
```

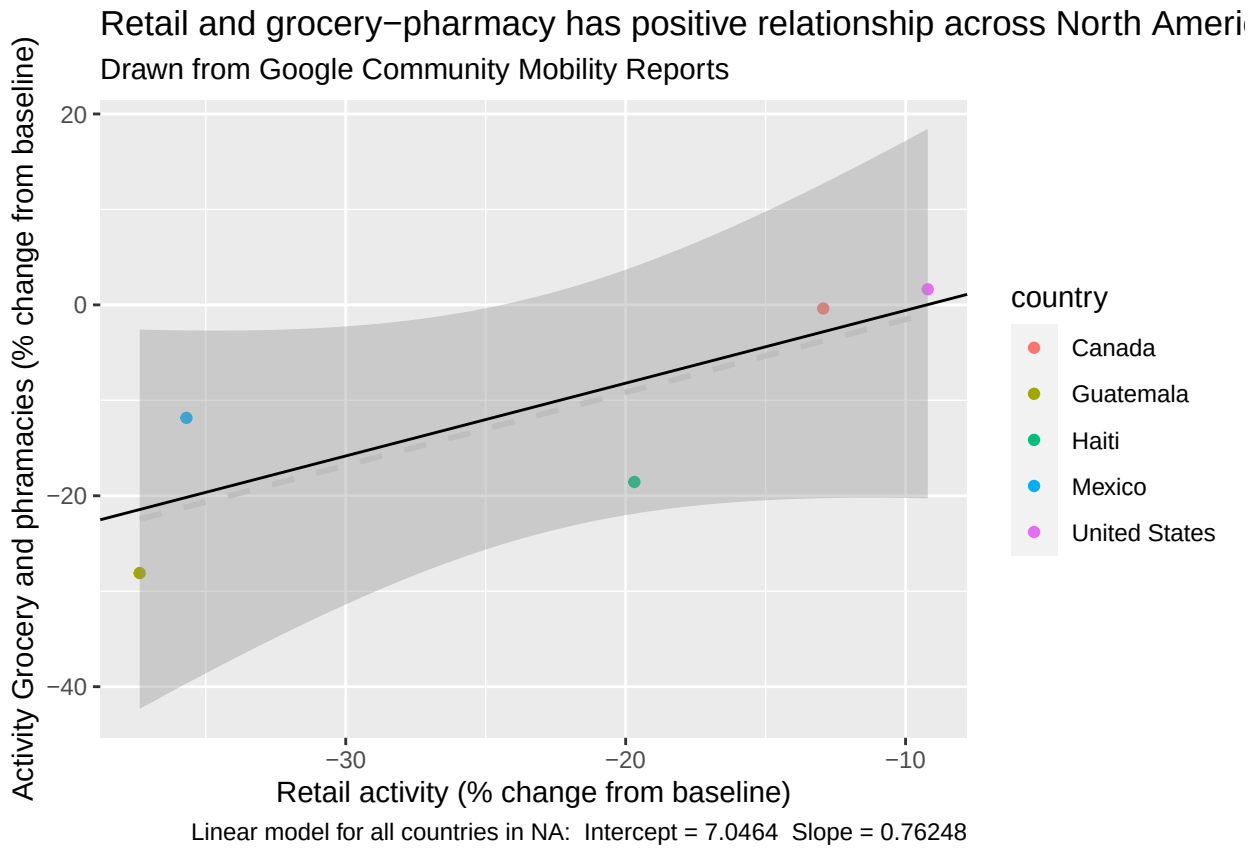
```
                      "Intercept =",signif(lm.EURhomework$coef[[1]],5),
                      " Slope =",signif(lm.EURhomework$coef[[2]],5)))
library(ggpubr)
ggarrange(EUR_home_work, legend="right")
```

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



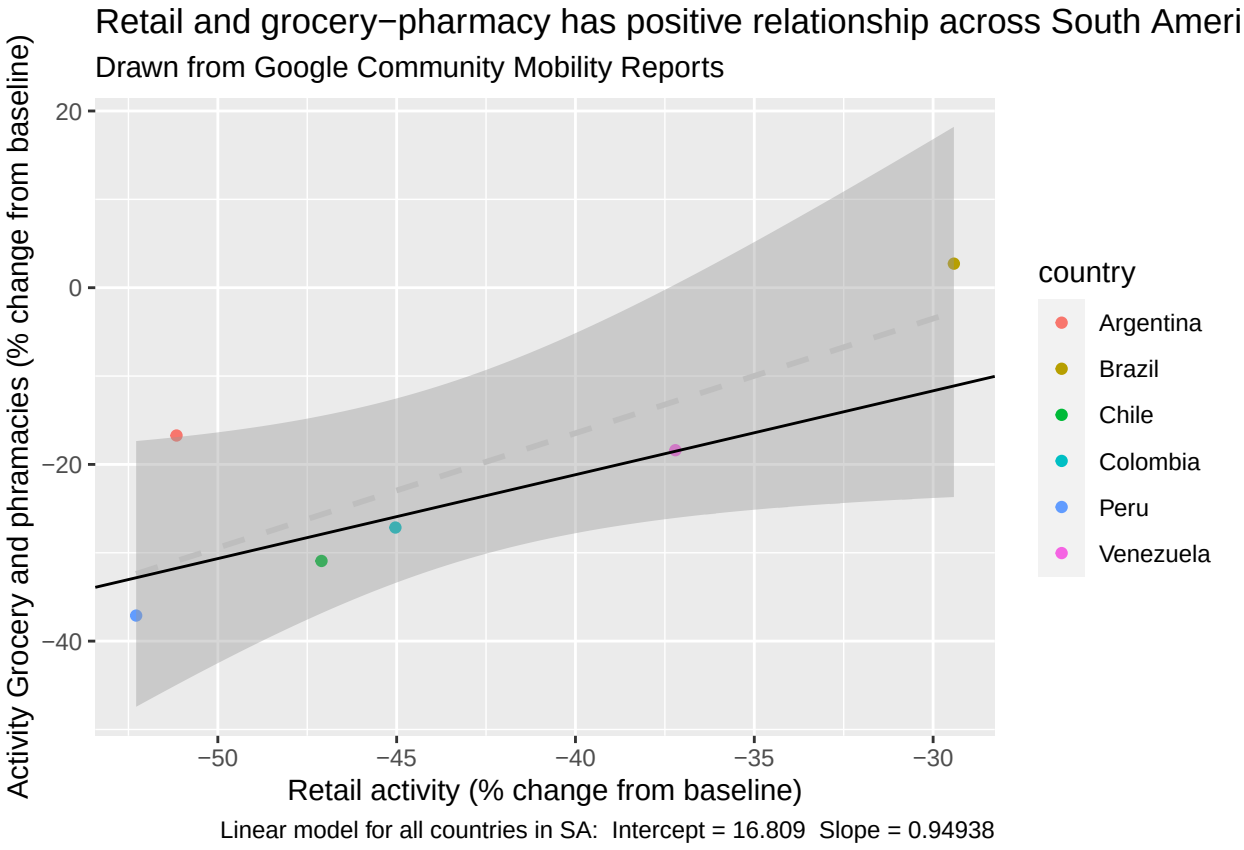

```
lm.NUSAhomework <- lm(mean_grocery_pharmacy ~ mean_retail, data = N.USA)
NUSA_home_work <- {N.USA1 %>% ggplot(aes(x = mean_retail,
                                         y = mean_grocery_pharmacy,
                                         color=country))} +
  xlab("Retail activity (% change from baseline)") +
  ylab("Activity Grocery and phramacies (% change from baseline)") +
  ggtitle("Retail and grocery-pharmacy has positive relationship across North America",
          subtitle = "Drawn from Google Community Mobility Reports") +
  geom_point() +
  stat_smooth(method = "lm", color="grey", linetype="dashed") +
  geom_abline(intercept = signif(lm.NUSAhomework$coef[[1]],5),
              slope = signif(lm.NUSAhomework$coef[[2]],5), color="black") +
  labs(caption = paste("Linear model for all countries in NA: ",
                       "Intercept =",signif(lm.NUSAhomework$coef[[1]],5),
                       " Slope =",signif(lm.NUSAhomework$coef[[2]],5)))
ggarrange(NUSA_home_work, legend="right")
```

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




```
lm.SUSAhomework <- lm(mean_grocery_pharmacy ~ mean_retail, data = S.USA)
SUSA_home_work <- {S.USA1 %>% ggplot(aes(x = mean_retail,
                                         y = mean_grocery_pharmacy,
                                         color=country))} +
  xlab("Retail activity (% change from baseline)") +
  ylab("Activity Grocery and phramacies (% change from baseline)") +
  ggtitle("Retail and grocery-pharmacy has positive relationship across South America",
          subtitle = "Drawn from Google Community Mobility Reports") +
  geom_point() +
  stat_smooth(method = "lm", color="grey", linetype="dashed") +
  geom_abline(intercept = signif(lm.SUSAhomework$coef[[1]],5),
              slope = signif(lm.SUSAhomework$coef[[2]],5), color="black") +
  labs(caption = paste("Linear model for all countries in SA: ",
                       "Intercept =",signif(lm.SUSAhomework$coef[[1]],5),
                       " Slope =",signif(lm.SUSAhomework$coef[[2]],5)))
ggarrange(SUSA_home_work, legend="right")
```

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


Retail and grocery−pharmacy has positive relationship across South Ameri
Drawn from Google Community Mobility Reports
Activity Grocery and phramacies (% change from baseline)
20
0
−20
−40
−50
−45
−40
−35
−30
Retail activity (% change from baseline)
Linear model for all countries in SA: Intercept = 16.809 Slope = 0.94938
country
Argentina
Brazil
Chile
Colombia
Peru
Venezuela