Gini Coefficients and GDP Growth

Setup

If you have installed the package "tidyverse", you do not need to run the next code chunk.

```
install.packages("tidyverse")
```

Although "readxl" and "broom" are part of the package "tidyverse', you need to load separately as they are not core and not loaded with 'tidyverse'.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## 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(readxl) # a package for Excel files, a part of tidyverse but not core
library(broom) # a package for modeling, a part of tidyverse but not core
library(WDI)
```

Importing

For Excel, you need to specify wich sheet you want to import by name as below or the number. You also need to skip several rows when blank rows are included.

```
df_gini_gdpgrowth_orig <- read_excel("data/P_Data_Extract_From_World_Development_Indicators 20 years db
    col_types = c("text", "text", "text",
        "text", "numeric", "numeric", "numeric",
        "numeric", "numeric", "numeric",
        "numeric", "numeric", "numeric",
        "numeric", "numeric", "numeric",
        "numeric", "numeric", "numeric",
        "numeric", "numeric"), na = "\t ..", n_max = 532)
```

Check the data frames.

```
df_gini_gdpgrowth_orig |> head()
```

```
## # A tibble: 6 x 24
##   `Country Name` `Country Code` `Series Name`      `Series Code` `2003 [YR2003]`
##   <chr>          <chr>          <chr>              <chr>                   <dbl>
## 1 Afghanistan    AFG            Gini index         SI.POV.GINI            NA
```

```
## 2 Afghanistan    AFG            GDP growth (annua~ NY.GDP.MKTP.~            8.83
## 3 Albania        ALB            Gini index         SI.POV.GINI             NA
## 4 Albania        ALB            GDP growth (annua~ NY.GDP.MKTP.~            5.53
## 5 Algeria        DZA            Gini index         SI.POV.GINI             NA
## 6 Algeria        DZA            GDP growth (annua~ NY.GDP.MKTP.~            7.20
## # i 19 more variables: `2004 [YR2004]` <dbl>, `2005 [YR2005]` <dbl>,
## #   `2006 [YR2006]` <dbl>, `2007 [YR2007]` <dbl>, `2008 [YR2008]` <dbl>,
## #   `2009 [YR2009]` <dbl>, `2010 [YR2010]` <dbl>, `2011 [YR2011]` <dbl>,
## #   `2012 [YR2012]` <dbl>, `2013 [YR2013]` <dbl>, `2014 [YR2014]` <dbl>,
## #   `2015 [YR2015]` <dbl>, `2016 [YR2016]` <dbl>, `2017 [YR2017]` <dbl>,
## #   `2018 [YR2018]` <dbl>, `2019 [YR2019]` <dbl>, `2020 [YR2020]` <dbl>,
## #   `2021 [YR2021]` <dbl>, `2022 [YR2022]` <dbl>
```

```
df_gini_gdpgrowth2 <- WDI(indicator = c(gini = "SI.POV.GINI", gdpgrowth = "NY.GDP.MKTP.KD.ZG"))
df_gini_gdpgrowth2 |> head()
```

```
##       country iso2c iso3c year gini gdpgrowth
## 1 Afghanistan    AF   AFG 1960   NA        NA
## 2 Afghanistan    AF   AFG 1961   NA        NA
## 3 Afghanistan    AF   AFG 1962   NA        NA
## 4 Afghanistan    AF   AFG 1963   NA        NA
## 5 Afghanistan    AF   AFG 1964   NA        NA
## 6 Afghanistan    AF   AFG 1965   NA        NA
```

Tidying

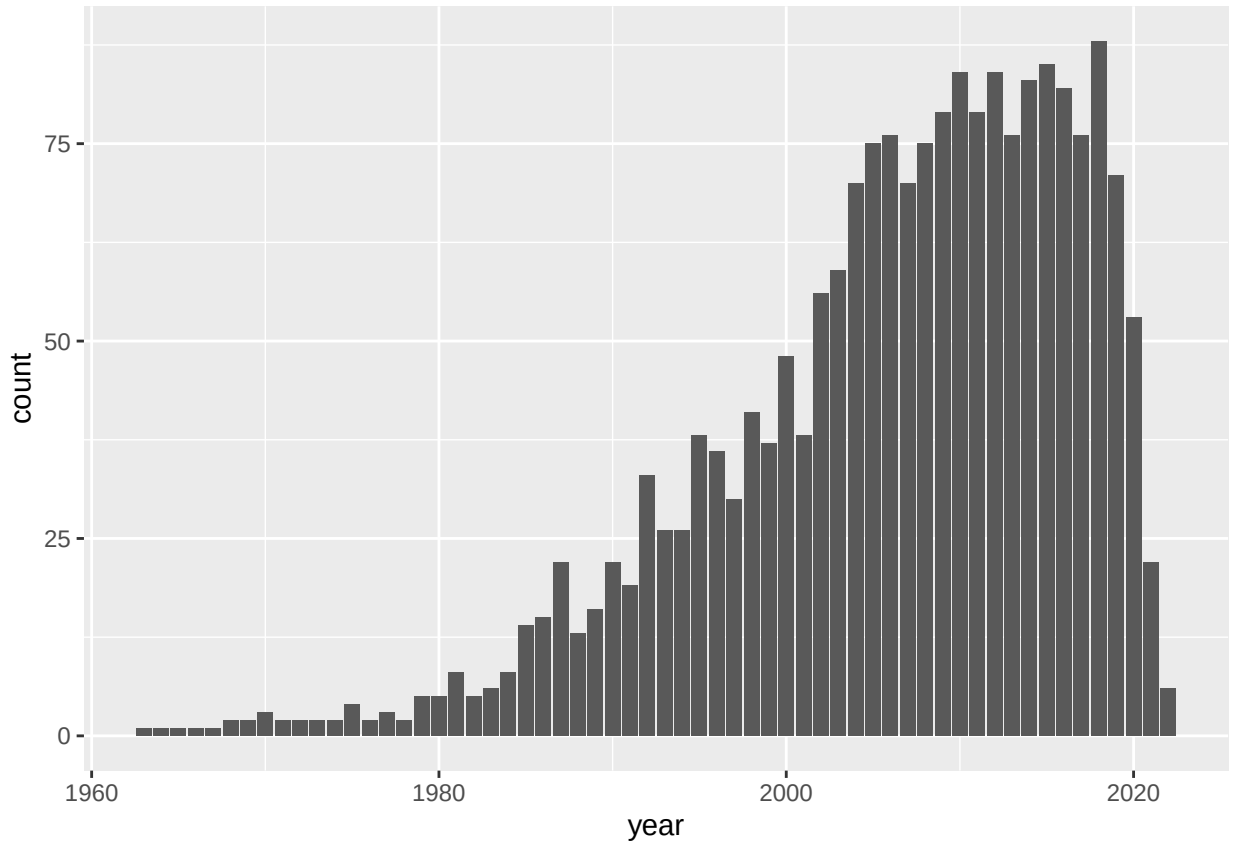
After tidying the original data, we obtained about the same as the one obtained above using WDI package. So I recommend you to get the data using WDI.

```
df_gini_gdpgrowth <- df_gini_gdpgrowth_orig |> pivot_longer(5:24, names_to = "year", values_to = "value
  mutate(year = as.numeric(str_sub(year, 1, 4))) |> select(-4) |>
  pivot_wider(names_from = `Series Name`, values_from = value)
df_gini_gdpgrowth |> head()
```

```
## # A tibble: 6 x 5
##   `Country Name` `Country Code`  year `Gini index` `GDP growth (annual %)`
##   <chr>          <chr>          <dbl>        <dbl>                   <dbl>
## 1 Afghanistan    AFG             2003           NA                    8.83
## 2 Afghanistan    AFG             2004           NA                    1.41
## 3 Afghanistan    AFG             2005           NA                   11.2
## 4 Afghanistan    AFG             2006           NA                    5.36
## 5 Afghanistan    AFG             2007           NA                   13.8
## 6 Afghanistan    AFG             2008           NA                    3.92
```

Number of available data

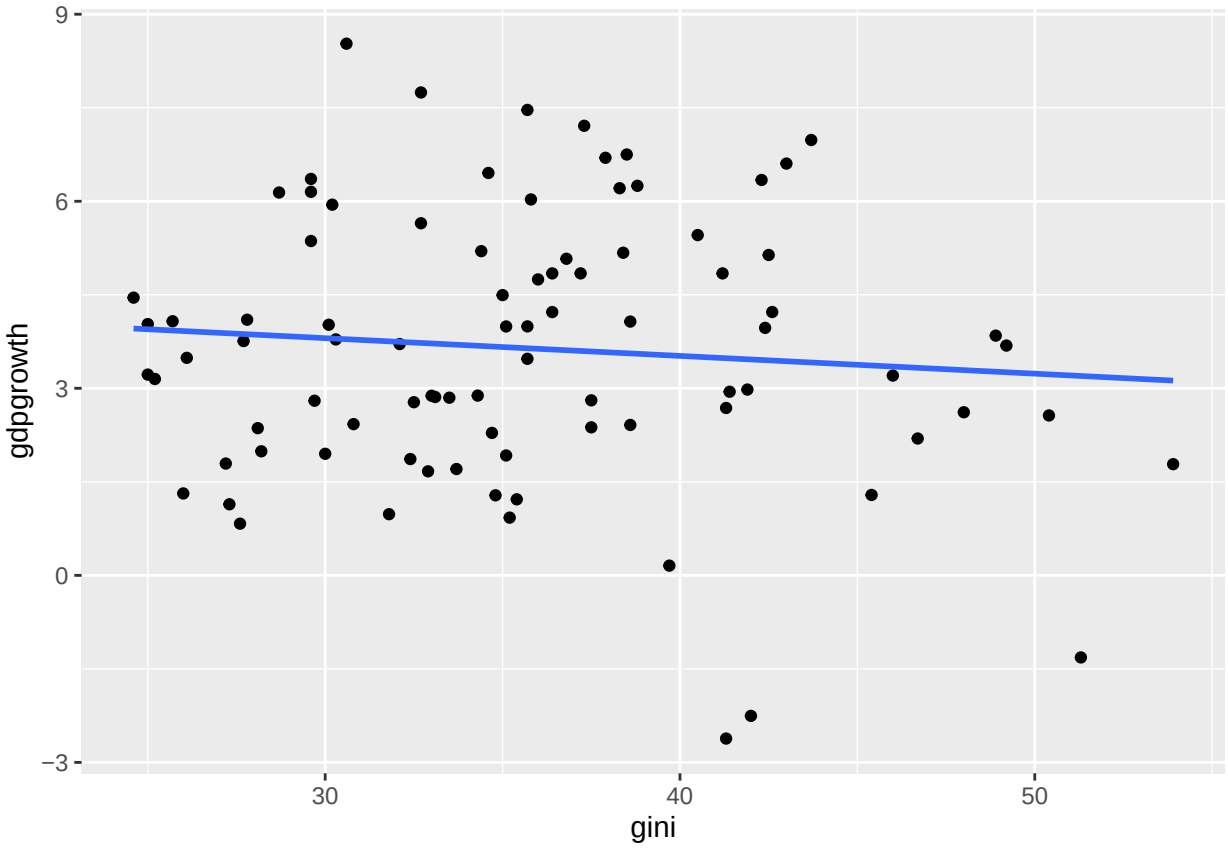
```
df_gini_gdpgrowth2 |> drop_na(gini, gdpgrowth) |>
  ggplot(aes(year)) + geom_bar()
```



Scatter Plots

It is better to fix a year.

```
YEAR <- 2018
df_gini_gdpgrowth2 |> filter (year == YEAR) |> drop_na(gini, gdpgrowth) |>
  ggplot(aes(gini, gdpgrowth)) + geom_point() + geom_smooth(method = "lm", formula = y~x, se = FALSE)
```



Models

```
YEAR <- 2018
model0 <- df_gini_gdpgrowth2 |> filter (year == YEAR) |> drop_na(gini, gdpgrowth) |> lm(gdpgrowth ~ gini
```

```
model0 |> summary()
```

```
##
## Call:
## lm(formula = gdpgrowth ~ gini, data = drop_na(filter(df_gini_gdpgrowth2,
##     year == YEAR), gini, gdpgrowth))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.1000 -1.3442 -0.0737  1.5272  4.7402
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.65888    1.25244   3.720 0.000355 ***
## gini        -0.02848    0.03448  -0.826 0.411083
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.154 on 86 degrees of freedom
## Multiple R-squared:  0.007871,   Adjusted R-squared:  -0.003665
## F-statistic: 0.6823 on 1 and 86 DF,  p-value: 0.4111
```

```
model0 |> summary() |> glance()
```

```
## # A tibble: 1 x 8
##   r.squared adj.r.squared sigma statistic p.value    df df.residual  nobs
##       <dbl>         <dbl> <dbl>     <dbl>   <dbl> <dbl>       <int> <dbl>
## 1   0.00787      -0.00367  2.15     0.682   0.411     1          86    88
```

```
model0 |> summary() |> tidy()
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)   4.66      1.25       3.72  0.000355
## 2 gini         -0.0285    0.0345    -0.826 0.411
```

```
model0 |> augment() |> head()
```

```
## # A tibble: 6 x 8
##   gdpgrowth  gini .fitted .resid   .hat .sigma  .cooksd .std.resid
##       <dbl> <dbl>   <dbl>  <dbl>  <dbl>  <dbl>    <dbl>      <dbl>
## 1      4.02  30.1    3.80  0.218 0.0194   2.17 0.000103      0.102
## 2     -1.32  51.3    3.20 -4.51  0.0737   2.11 0.189        -2.18
## 3     -2.62  41.3    3.48 -6.10  0.0194   2.06 0.0808       -2.86
## 4      5.20  34.4    3.68  1.52  0.0118   2.16 0.00301       0.710
## 5      2.88  34.3    3.68 -0.799 0.0119   2.16 0.000837     -0.373
## 6      2.43  30.8    3.78 -1.36  0.0175   2.16 0.00360      -0.635
```