Access to Energy

Deepak Dubey

21/10/2021

Setting up my environment

Notes: setting up the R environment by loading the 'tidyverse' and the other packages

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

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
install.packages("here")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
library(here)
```

```
## here() starts at /cloud/project
```

```
library(lubridate)
```

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

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

```
library(dplyr)
install.packages("skimr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
library(skimr)
install.packages("janitor")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
library(janitor)
```

```
##
## Attaching package: 'janitor'
```

```
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(ggplot2)
```

Collect the data and save it as a new data frame

```
read_csv("number-of-people-with-and-without-electricity-access.csv")
```

```
## Rows: 6958 Columns: 5
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Entity, Code
## dbl (3): Year, Number of people with access to electricity, Number of people...
```

```
##
## 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.
```

```
## # A tibble: 6,958 x 5
##    Entity      Code   Year `Number of people with ac~ `Number of people without~
##    <chr>       <chr> <dbl>                       <dbl>                      <dbl>
##  1 Afghanistan AFG    1990                        1225                   12247889
##  2 Afghanistan AFG    1991                        1299                   12992358
##  3 Afghanistan AFG    1992                        1398                   13979833
##  4 Afghanistan AFG    1993                        1510                   15093589
##  5 Afghanistan AFG    1994                        1617                   16171102
##  6 Afghanistan AFG    1995                        1710                   17097831
##  7 Afghanistan AFG    1996                        1782                   17821102
##  8 Afghanistan AFG    1997                        1838                   18379767
##  9 Afghanistan AFG    1998                        4146                   18859853
## 10 Afghanistan AFG    1999                       34856                   19368820
## # ... with 6,948 more rows
```

```
p <- read_csv("number-of-people-with-and-without-electricity-access.csv")
```

```
## Rows: 6958 Columns: 5
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Entity, Code
## dbl (3): Year, Number of people with access to electricity, Number of people...
```

```
##
## 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.
```

Clean up and wrangle the data

```
str(p)
```

```
## spec_tbl_df [6,958 x 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Entity                                          : chr [1:6958] "Afghanistan" "Afghanistan" "Afghanis
##  $ Code                                            : chr [1:6958] "AFG" "AFG" "AFG" "AFG" ...
##  $ Year                                            : num [1:6958] 1990 1991 1992 1993 1994 ...
##  $ Number of people with access to electricity     : num [1:6958] 1225 1299 1398 1510 1617 ...
##  $ Number of people without access to electricity: num [1:6958] 12247889 12992358 13979833 15093589 
##  - attr(*, "spec")=
##   .. cols(
##   ..   Entity = col_character(),
##   ..   Code = col_character(),
##   ..   Year = col_double(),
##   ..   `Number of people with access to electricity` = col_double(),
##   ..   `Number of people without access to electricity` = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
skim_without_charts(p)
```

Table 1: Data summary

Name	p
Number of rows	6958
Number of columns	5
Column type frequency:	
character	2
numeric	3
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Entity	0	1.00	4	50	0	261	0
Code	1242	0.82	3	8	0	215	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Year	0	1	2003.1	7.77	1990	1996.0	2003.0	2010	2016
Number of people with access to electricity	0	1	200035721	663047668.31	0	478721.2	4875067.5	44366620	1504588805
Number of people without access to electricity	0	1	58424469.4	195884149.38	0	0.0	263609.5	8858421	1526744598

```
head(p)
```

```
## # A tibble: 6 x 5
##   Entity      Code   Year `Number of people with ac~ `Number of people without ~
##   <chr>       <chr> <dbl>                       <dbl>                        <dbl>
## 1 Afghanistan AFG    1990                        1225                     12247889
## 2 Afghanistan AFG    1991                        1299                     12992358
## 3 Afghanistan AFG    1992                        1398                     13979833
## 4 Afghanistan AFG    1993                        1510                     15093589
## 5 Afghanistan AFG    1994                        1617                     16171102
## 6 Afghanistan AFG    1995                        1710                     17097831
```

```
glimpse(p)
```

```
## Rows: 6,958
## Columns: 5
## $ Entity                                          <chr> "Afghanistan", "Afgha~
## $ Code                                            <chr> "AFG", "AFG", "AFG", ~
## $ Year                                            <dbl> 1990, 1991, 1992, 199~
## $ `Number of people with access to electricity`   <dbl> 1225, 1299, 1398, 151~
## $ `Number of people without access to electricity` <dbl> 12247889, 12992358, 1~
```

```
colnames(p)
```

```
## [1] "Entity"
## [2] "Code"
## [3] "Year"
## [4] "Number of people with access to electricity"
## [5] "Number of people without access to electricity"
```

```
clean_names(p)
```

```
## # A tibble: 6,958 x 5
##    entity      code   year number_of_people_with_acc~ number_of_people_without_~
##    <chr>       <chr> <dbl>                       <dbl>                        <dbl>
##  1 Afghanistan AFG    1990                        1225                     12247889
##  2 Afghanistan AFG    1991                        1299                     12992358
##  3 Afghanistan AFG    1992                        1398                     13979833
##  4 Afghanistan AFG    1993                        1510                     15093589
##  5 Afghanistan AFG    1994                        1617                     16171102
##  6 Afghanistan AFG    1995                        1710                     17097831
##  7 Afghanistan AFG    1996                        1782                     17821102
##  8 Afghanistan AFG    1997                        1838                     18379767
##  9 Afghanistan AFG    1998                        4146                     18859853
## 10 Afghanistan AFG    1999                       34856                     19368820
## # ... with 6,948 more rows
```

```
p <- rename(p,Country_name=Entity)
```

Add data to prepare for Analysis, and conduct for descriptive analysis

```
colnames(p)
```

```
## [1] "Country_name"
## [2] "Code"
## [3] "Year"
```

```
## [4] "Number of people with access to electricity"
## [5] "Number of people without access to electricity"
```

```
nrow(p)
```

```
## [1] 6958
```

```
dim(p)
```

```
## [1] 6958    5
```

```
head(p)
```

```
## # A tibble: 6 x 5
##   Country_name Code   Year `Number of people with ac~ `Number of people without~
##   <chr>        <chr> <dbl>                      <dbl>                      <dbl>
## 1 Afghanistan  AFG    1990                       1225                   12247889
## 2 Afghanistan  AFG    1991                       1299                   12992358
## 3 Afghanistan  AFG    1992                       1398                   13979833
## 4 Afghanistan  AFG    1993                       1510                   15093589
## 5 Afghanistan  AFG    1994                       1617                   16171102
## 6 Afghanistan  AFG    1995                       1710                   17097831
```

```
str(p)
```

```
## spec_tbl_df [6,958 x 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Country_name                                  : chr [1:6958] "Afghanistan" "Afghanistan" "Afghanis
##  $ Code                                          : chr [1:6958] "AFG" "AFG" "AFG" "AFG" ...
##  $ Year                                          : num [1:6958] 1990 1991 1992 1993 1994 ...
##  $ Number of people with access to electricity   : num [1:6958] 1225 1299 1398 1510 1617 ...
##  $ Number of people without access to electricity: num [1:6958] 12247889 12992358 13979833 15093589 1
##  - attr(*, "spec")=
##   .. cols(
##   ..   Entity = col_character(),
##   ..   Code = col_character(),
##   ..   Year = col_double(),
##   ..   `Number of people with access to electricity` = col_double(),
##   ..   `Number of people without access to electricity` = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
summary(p)
```

```
##  Country_name           Code                Year
##  Length:6958        Length:6958        Min.   :1990
##  Class :character   Class :character   1st Qu.:1996
##  Mode  :character   Mode  :character   Median :2003
##                                        Mean   :2003
##                                        3rd Qu.:2010
##                                        Max.   :2016
##  Number of people with access to electricity
##  Min.   :0.000e+00
##  1st Qu.:4.787e+05
##  Median :4.875e+06
##  Mean   :2.000e+08
##  3rd Qu.:4.437e+07
##  Max.   :6.505e+09
##  Number of people without access to electricity
```

```
##  Min.   :0.000e+00
##  1st Qu.:0.000e+00
##  Median :2.636e+05
##  Mean   :5.842e+07
##  3rd Qu.:8.858e+06
##  Max.   :1.527e+09
```

Organising the data

```
p %>% select(-`Number of people with access to electricity`)
```

```
## # A tibble: 6,958 x 4
##    Country_name Code   Year `Number of people without access to electricity`
##    <chr>        <chr> <dbl>                                            <dbl>
##  1 Afghanistan  AFG    1990                                         12247889
##  2 Afghanistan  AFG    1991                                         12992358
##  3 Afghanistan  AFG    1992                                         13979833
##  4 Afghanistan  AFG    1993                                         15093589
##  5 Afghanistan  AFG    1994                                         16171102
##  6 Afghanistan  AFG    1995                                         17097831
##  7 Afghanistan  AFG    1996                                         17821102
##  8 Afghanistan  AFG    1997                                         18379767
##  9 Afghanistan  AFG    1998                                         18859853
## 10 Afghanistan  AFG    1999                                         19368820
## # ... with 6,948 more rows
```

```
p %>% arrange(Country_name)
```

```
## # A tibble: 6,958 x 5
##    Country_name Code   Year `Number of people with ac~ `Number of people withou~
##    <chr>        <chr> <dbl>                      <dbl>                      <dbl>
##  1 Afghanistan  AFG    1990                       1225                   12247889
##  2 Afghanistan  AFG    1991                       1299                   12992358
##  3 Afghanistan  AFG    1992                       1398                   13979833
##  4 Afghanistan  AFG    1993                       1510                   15093589
##  5 Afghanistan  AFG    1994                       1617                   16171102
##  6 Afghanistan  AFG    1995                       1710                   17097831
##  7 Afghanistan  AFG    1996                       1782                   17821102
##  8 Afghanistan  AFG    1997                       1838                   18379767
##  9 Afghanistan  AFG    1998                       4146                   18859853
## 10 Afghanistan  AFG    1999                      34856                   19368820
## # ... with 6,948 more rows
```

```
p %>% arrange(-Year)
```

```
## # A tibble: 6,958 x 5
##    Country_name        Code   Year `Number of people wit~ `Number of people wit~
##    <chr>               <chr> <dbl>                  <dbl>                  <dbl>
##  1 Afghanistan         AFG    2016               29158594                5497438
##  2 Albania             ALB    2016                2876101                      0
##  3 Algeria             DZA    2016               40378483                 227569
##  4 Andorra             AND    2016                  77281                      0
##  5 Angola              AGO    2016               11675390               17138073
##  6 Antigua and Barbuda ATG    2016                  98292                   2671
##  7 Arab World          <NA>   2016              360802582               45650108
```

```
##  8 Argentina           ARG    2016               43847430                     0
##  9 Armenia             ARM    2016                2924816                     0
## 10 Aruba               ABW    2016                 100179                  4643
## # ... with 6,948 more rows
```

```
p2 <- p %>% arrange(-Year)
p %>% arrange(-`Number of people without access to electricity`)
```

```
## # A tibble: 6,958 x 5
##    Country_name        Code      Year `Number of people wi~ `Number of people w~
##    <chr>               <chr>    <dbl>                 <dbl>                <dbl>
##  1 World               OWID_WRL  1991            3848744021           1526744598
##  2 World               OWID_WRL  1990            3775207395           1512895819
##  3 Low & middle income <NA>      1991            2832587548           1497101663
##  4 World               OWID_WRL  1992            3962785379           1496968486
##  5 World               OWID_WRL  1994            4138408267           1490382909
##  6 World               OWID_WRL  1995            4229597492           1484196880
##  7 World               OWID_WRL  1993            4060806959           1484066129
##  8 Low & middle income <NA>      1990            2767768994           1482999753
##  9 IDA & IBRD total    <NA>      1991            2912376176           1480409806
## 10 Low & middle income <NA>      1992            2939106996           1467990058
## # ... with 6,948 more rows
```

```
p3 <- p %>% arrange(-`Number of people without access to electricity`)
drop_na(p)
```

```
## # A tibble: 5,716 x 5
##    Country_name Code   Year `Number of people with ac~ `Number of people withou~
##    <chr>        <chr> <dbl>                      <dbl>                     <dbl>
##  1 Afghanistan  AFG    1990                       1225                  12247889
##  2 Afghanistan  AFG    1991                       1299                  12992358
##  3 Afghanistan  AFG    1992                       1398                  13979833
##  4 Afghanistan  AFG    1993                       1510                  15093589
##  5 Afghanistan  AFG    1994                       1617                  16171102
##  6 Afghanistan  AFG    1995                       1710                  17097831
##  7 Afghanistan  AFG    1996                       1782                  17821102
##  8 Afghanistan  AFG    1997                       1838                  18379767
##  9 Afghanistan  AFG    1998                       4146                  18859853
## 10 Afghanistan  AFG    1999                      34856                  19368820
## # ... with 5,706 more rows
```

```
p4 <- drop_na(p)
p %>% group_by(Country_name) %>% drop_na() %>% summarise(mean_Number_of_people_without_access_to_electr
```

```
## # A tibble: 215 x 2
##    Country_name        mean_Number_of_people_without_access_to_electricity
##    <chr>                                                             <dbl>
##  1 Afghanistan                                                    15634096.
##  2 Albania                                                              0
##  3 Algeria                                                        360750.
##  4 Andorra                                                              0
##  5 Angola                                                         13725326.
##  6 Antigua and Barbuda                                                7102.
##  7 Argentina                                                      1440683.
##  8 Armenia                                                          28794.
##  9 Aruba                                                             7070.
```

```
## 10 Australia                                                                0
## # ... with 205 more rows
```

```
p5 <- p %>% group_by(Country_name) %>% drop_na() %>% summarise(mean_Number_of_people_without_access_to_
p %>%
  group_by(Country_name) %>%
  drop_na() %>%
  summarise(max_Number_of_people_without_access_to_electricity = max(`Number of people without access t
```

```
## # A tibble: 215 x 2
##    Country_name        max_Number_of_people_without_access_to_electricity
##    <chr>                                                            <dbl>
##  1 Afghanistan                                                   20803651
##  2 Albania                                                             0
##  3 Algeria                                                        464066
##  4 Andorra                                                             0
##  5 Angola                                                        18305917
##  6 Antigua and Barbuda                                              9922
##  7 Argentina                                                     3063234
##  8 Armenia                                                         82072
##  9 Aruba                                                            7992
## 10 Australia                                                           0
## # ... with 205 more rows
```

```
p6 <- p %>%
  group_by(Country_name) %>%
  drop_na() %>%
  summarise(max_Number_of_people_without_access_to_electricity = max(`Number of people without access t
p %>% filter(Country_name == 'India', Year == '2016')
```

```
## # A tibble: 1 x 5
##   Country_name Code   Year `Number of people with ac~ `Number of people without~
##   <chr>        <chr> <dbl>                      <dbl>                      <dbl>
## 1 India        IND    2016                 1119279901                  204891453
```

```
p7 <- p %>% filter(Country_name == 'India')
```

Visualizations

Here, we will go through a series of visualizations - I've curated compelling and intuitive visuals in **Tableau**, under the name of the title 'How many people don't have access to electricity?'click herelink