Quality of Education in Indonesia in 2022

Group: 1
Name:

1. Aaron Winston Gho - 2702210522
2. Aldo Oktavianus - 2702234081
3. Kelvin Jonathan Yusach - 2702209533
4. Nicholas Ananda Heryanto - 2702213695
5. Tyrone Yutanesy Iman - 2702211512

Hello everyone! Today, we're diving into a crucial topic: the quality of education in Indonesia in 2022. Education is the cornerstone of any nation's progress and development, and it's especially significant for Indonesia, a country with a vast and diverse population.

In this video, we'll explore various aspects of the educational landscape in Indonesia, focusing on recent statistics and trends. But first, let's frame our discussion within a global context by linking it to the United Nations Sustainable Development Goal number 4.

SDG 4 aims to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all." These targets are designed to address the global challenge of educational inequality and to ensure that education systems are equipped to prepare students for the future. Indonesia, like many countries, faces unique challenges in achieving these goals. The nation's vast archipelago, diverse cultures, and varying levels of economic development across regions create a complex educational landscape.

Since the data for 2023 is not yet complete, we decided to use data from 2022. We are using data from the Badan Pusat Statistik (BPS) and Kemendikbud Ristek. In 2022, significant efforts have been made to improve educational access and quality, but there remain areas that need further attention and resources.

In the following segments, we'll present detailed insights into:

1. Distribution of teachers and how it impacts educational quality.
2. The trends in school participation rates over the past two decades from 2002 - 2022.
3. The relationship between educational attainment and poverty levels.
4. The distribution and accessibility of libraries across the country.

By examining these aspects, we aim to provide a comprehensive overview of the state of education in Indonesia and highlight areas where progress aligns with the objectives of SDG 4. Let's get started!

================ Data Cleaning================

1. Cleaning Data for Distribution of Teachers in Indonesia in 2022

```
library(dplyr)
data = read.table(file.choose(), sep=",", header = T, skip = 1)

# looking for missing data
missing_summary <- data %>%
  summarise(across(everything(), ~ sum(is.na(.))))
print(missing_summary)
```

```
##    Aceh X3336 X193 X3529 X45823 X2718 X48541 X451027 X35931 X486958
## 1     0     0    0     0       0     0       0       0       0       0
```

```
summary(data)
```

```
##       Aceh                X3336              X193             X3529            X45823
##  Length:34          Min.   :   437   Min.   :   29.0   Min.   :   485   Min.   :  5489
##  Class :character   1st Qu.:  1300   1st Qu.:  182.5   1st Qu.:  1821   1st Qu.: 13490
##  Mode  :character   Median :  2360   Median :  327.5   Median :  2645   Median : 24892
##                     Mean   :  7551   Mean   : 1108.0   Mean   :  8659   Mean   : 79432
##                     3rd Qu.:  4095   3rd Qu.:  829.0   3rd Qu.:  4680   3rd Qu.: 44641
##                     Max.   :130042   Max.   :18933.0   Max.   :148975   Max.   :1373257
##       X2718              X48541             X451027            X35931            X486958
##  Min.   :   268   Min.   :   6095   Min.   :   69072   Min.   :   4228   Min.   :   78529
##  1st Qu.:  2228   1st Qu.:  18176   1st Qu.:  158935   1st Qu.:  33530   1st Qu.:  217822
##  Median :  4466   Median :  27168   Median :  339414   Median :  65685   Median :  397546
##  Mean   : 13582   Mean   :  93014   Mean   : 1184745   Mean   : 217198   Mean   : 1401943
##  3rd Qu.:  8365   3rd Qu.:  56237   3rd Qu.:  698306   3rd Qu.: 172856   3rd Qu.:  793096
##  Max.   :232252   Max.   :1605509   Max.   :20366178   Max.   :3710333   Max.   :24076511
```

```
column_types <- sapply(data, class)

# replace data with "Unknown" if missing data is char
dataNew <- data %>%
  mutate_if(is.character, ~ifelse(is.na(.), "Unknown", .))

# replace data with column's median if missing data is numerical
dataNew <- data %>%
  mutate_all(
    funs(if(is.numeric(.)) replace(., is.na(.), median(., na.rm = TRUE)) else .)
  )
```

```
## Warning: 'funs()' was deprecated in dplyr 0.8.0.
## i Please use a list of either functions or lambdas:
##
## # Simple named list: list(mean = mean, median = median)
##
## # Auto named with 'tibble::lst()': tibble::lst(mean, median)
##
## # Using lambdas list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was generated.
```

```
# Looking for Duplicated Data
duplicated_rows <- duplicated(data)
duplicated_rows
```

```
## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [16] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [31] FALSE FALSE FALSE FALSE
```

```
# Make a new data without duplicates in variabel called unique data specific
unique_data_specific <- distinct(data, .keep_all = TRUE)
unique_data_specific
```

	Aceh	X3336	X193	X3529	X45823	X2718	X48541	X451027	X35931	X486958
1	Sumatera Utara	8205	1577	9782	91137	18448	109585	1204565	344087	1548652
2	Sumatera Barat	3962	298	4260	42372	4461	46833	531193	66025	597218
3	Riau	3221	570	3791	42764	8614	51378	634909	138617	773526
4	Jambi	2308	155	2463	24989	2172	27161	335196	33498	368694
5	Sumatera Selatan	4252	444	4696	54232	5629	59861	796801	85733	882534
6	Bengkulu	1290	108	1398	14351	1684	16035	173543	20838	194381
7	Lampung	4314	418	4732	52422	5289	57711	719438	80181	799619
8	Kepulauan Bangka Belitung	760	72	832	8672	1006	9678	146317	16405	162722
9	Kepulauan Riau	684	289	973	9483	4550	14033	154066	71509	225575
10	DKI Jakarta	1308	931	2239	24794	15550	40344	546713	218377	765090
11	Jawa Barat	17302	2357	19659	177306	35543	212849	3883339	567595	4450934
12	Jawa Tengah	17472	1288	18760	156270	17401	173671	2322013	309161	2631174
13	DI Yogyakarta	1423	425	1848	14788	5667	20455	190728	83710	274438
14	Jawa Timur	16980	2027	19007	161954	24844	186798	2184504	395409	2579913
15	Banten	3895	739	4634	45266	11621	56887	996698	189185	1185883
16	Bali	2265	151	2416	23162	2557	25719	343632	39232	382864
17	Nusa Tenggara Barat	3011	298	3309	39286	2873	42159	472110	37359	509469
18	Nusa Tenggara Timur	3378	1825	5203	36362	17923	54285	417555	251670	669225
19	Kalimantan Barat	4139	299	4438	36209	3331	39540	504176	61472	565648
20	Kalimantan Tengah	2412	236	2648	22934	2551	25485	232282	42758	275040
21	Kalimantan Selatan	2740	175	2915	27785	2397	30182	323293	33626	356919
22	Kalimantan Timur	1652	267	1919	22416	3696	26112	346884	65345	412229
23	Kalimantan Utara	437	48	485	5489	606	6095	69072	9457	78529
24	Sulawesi Utara	1344	859	2203	12705	6090	18795	132584	82653	215237
25	Sulawesi Tengah	2670	256	2926	25032	2143	27175	281651	26071	307722
26	Sulawesi Selatan	6082	356	6438	65702	4702	70404	811242	68935	880177
27	Sulawesi Tenggara	2257	88	2345	24705	862	25567	279295	12428	291723
28	Gorontalo	896	32	928	8108	284	8392	103914	4228	108142
29	Sulawesi Barat	1297	29	1326	13487	268	13755	139578	4415	143993
30	Maluku	1276	536	1812	13498	4471	17969	148988	59008	207996

```
## 31              Maluku  Utara   1110   205   1315  10439  1598  12037  117634  22930  140564
## 32              Papua   Barat    700   404   1104   6705  3086   9791   82955  48216  131171
## 33              Papua           1664   978   2642  12610  7617  20227  288283 184269  472552
## 34           Indonesia  130042 18933 148975 1373257 232252 1605509 20366178 3710333 24076511
```

```r
# Function to detect outliers using IQR method
outliers_iqr <- function(df) {
  df %>%
    summarise(across(where(is.numeric), ~{
      Q1 <- quantile(., 0.25, na.rm = TRUE)
      Q3 <- quantile(., 0.75, na.rm = TRUE)
      IQR <- Q3 - Q1
      lower_bound <- Q1 - 1.5 * IQR
      upper_bound <- Q3 + 1.5 * IQR
      sum(. < lower_bound | . > upper_bound, na.rm = TRUE)
    }, .names = "IQR_{.col}"))
}

# Function to detect outliers using Z-Score method
outliers_zscore <- function(df, threshold = 3) {
  df %>%
    summarise(across(where(is.numeric), ~{
      z_scores <- scale(.)
      sum(abs(z_scores) > threshold, na.rm = TRUE)
    }, .names = "ZScore_{.col}"))
}

# Detecting outliers
iqr_outliers_summary <- outliers_iqr(data)
zscore_outliers_summary <- outliers_zscore(data)

# Print
print(iqr_outliers_summary)
```

```
##   IQR_X3336 IQR_X193 IQR_X3529 IQR_X45823 IQR_X2718 IQR_X48541 IQR_X451027 IQR_X35931 IQR_X486958
## 1         4        4         5          4         5          4           4          3           4
```

```r
print(zscore_outliers_summary)
```

```
##   ZScore_X3336 ZScore_X193 ZScore_X3529 ZScore_X45823 ZScore_X2718 ZScore_X48541 ZScore_X451027
## 1            1           1           1             1           1           1             1
##   ZScore_X35931 ZScore_X486958
## 1             1              1
```

Overview and Importing Datasets

Here's an explanation on our group's coding cleaning process. First of all we will access the dplyr and skimr libraries. After that we will read the file from the data we provide. Here there are 4 data, we take 4 factors, namely the number of libraries, the number of school participation per age level, the number of schools and teachers, and finally the influence of poverty on completing education. First of all, of course, we will overview the dataframe which will be reviewed with a view and see the summary data with a skim.

Detect Missing data

After that, we will see the amount of missing data in our file. Here the system will detect and count the number of missing data per column via summarise(across(everything(), ~ sum(is.na(.)))).

Changing Missing Data

After detecting missing data, we will change the missing data. If the data variable is in the form of characters and there is missing data, it will be changed to "Unknown", this aims to not eliminate the uniqueness of the data because if we replace it with the most mode then we have removed the uniqueness of the data. Meanwhile, we will change the numerical data by changing the value of the missing data to the median. This aims to prevent skewness in the data so that our data maintains its distribution.

Duplicated Data and Outliers

Next, we will look for whether there is duplicated data. If there is duplicated data, we will store our data in a new dataframe that does not have duplicated data. After cleaning missing and duplicated data, we clean outliers contained in the data. Here we use two methods at once to detect outliers using the IQR method and Z-score. Here we create an IQR and Z-score function that accepts data frame parameters so that we can easily know whether our data has outliers or not.

2. Cleaning Data for School Participation Rates in Indonesia from 2002 to 2022

```
library(dplyr)
data = read_excel(file.choose())

# looking for missing data
missing_summary <- data %>%
  summarise(across(everything(), ~ sum(is.na(.))))
print(missing_summary)
```

```
## # A tibble: 1 x 5
##   Tahun `7-12` `13-15` `16-18` `19-24`
##   <int>  <int>   <int>   <int>   <int>
## 1     0      0       0       0       0
```

```
summary(data)
```

```
##      Tahun          7-12           13-15           16-18           19-24      
##  Min.   :2002   Min.   :96.10   Min.   :79.21   Min.   :49.76   Min.   :11.38  
##  1st Qu.:2007   1st Qu.:97.58   1st Qu.:84.65   1st Qu.:55.16   1st Qu.:12.72  
##  Median :2012   Median :98.02   Median :89.76   Median :61.42   Median :16.13  
##  Mean   :2012   Mean   :98.17   Mean   :89.48   Mean   :62.56   Mean   :18.29  
##  3rd Qu.:2017   3rd Qu.:99.10   3rd Qu.:95.08   3rd Qu.:71.42   3rd Qu.:24.40  
##  Max.   :2022   Max.   :99.26   Max.   :95.99   Max.   :73.15   Max.   :26.01
```

```
column_types <- sapply(data, class)

# replace data with "Unknown" if missing data is char
dataNew <- data %>%
  mutate_if(is.character, ~ifelse(is.na(.), "Unknown", .))
```

```r
# replace data with column's median if missing data is numerical
dataNew <- data %>%
  mutate_all(
    funs(if(is.numeric(.)) replace(., is.na(.), median(., na.rm = TRUE)) else .)
  )
```

```
## Warning: 'funs()' was deprecated in dplyr 0.8.0.
## i Please use a list of either functions or lambdas:
##
## # Simple named list: list(mean = mean, median = median)
##
## # Auto named with 'tibble::lst()': tibble::lst(mean, median)
##
## # Using lambdas list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was generated.
```

```r
# Looking for Duplicated Data
duplicated_rows <- duplicated(data)
duplicated_rows
```

```
##  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [16] FALSE FALSE FALSE FALSE FALSE FALSE
```

```r
# Make a new data without duplicates in variabel called unique data specific
unique_data_specific <- distinct(data, .keep_all = TRUE)
unique_data_specific
```

```
## # A tibble: 21 x 5
##    Tahun '7-12' '13-15' '16-18' '19-24'
##    <dbl>  <dbl>   <dbl>   <dbl>   <dbl>
##  1  2002   96.1    79.2    49.8    11.6
##  2  2003   96.4    81.0    51.0    11.7
##  3  2004   96.8    83.5    53.5    12.1
##  4  2005   97.1    84.0    53.9    12.2
##  5  2006   97.4    84.1    53.9    11.4
##  6  2007   97.6    84.6    55.5    13.1
##  7  2008   97.9    84.9    55.5    13.3
##  8  2009   98.0    85.5    55.2    12.7
##  9  2010   98.0    86.2    56.0    13.8
## 10  2011   97.6    87.8    57.8    14.3
## # i 11 more rows
```

```r
# Function to detect outliers using IQR method
outliers_iqr <- function(df) {
  df %>%
    summarise(across(where(is.numeric), ~{
      Q1 <- quantile(., 0.25, na.rm = TRUE)
      Q3 <- quantile(., 0.75, na.rm = TRUE)
      IQR <- Q3 - Q1
      lower_bound <- Q1 - 1.5 * IQR
      upper_bound <- Q3 + 1.5 * IQR
```

```
      sum(. < lower_bound | . > upper_bound, na.rm = TRUE)
    }, .names = "IQR_{.col}"))
}

# Function to detect outliers using Z-Score method
outliers_zscore <- function(df, threshold = 3) {
  df %>%
    summarise(across(where(is.numeric), ~{
      z_scores <- scale(.)
      sum(abs(z_scores) > threshold, na.rm = TRUE)
    }, .names = "ZScore_{.col}"))
}

# Detecting outliers
iqr_outliers_summary <- outliers_iqr(data)
zscore_outliers_summary <- outliers_zscore(data)

# Print
print(iqr_outliers_summary)
```

```
## # A tibble: 1 x 5
##   IQR_Tahun `IQR_7-12` `IQR_13-15` `IQR_16-18` `IQR_19-24`
##       <int>       <int>       <int>       <int>       <int>
## 1         0           0           0           0           0
```

```
print(zscore_outliers_summary)
```

```
## # A tibble: 1 x 5
##   ZScore_Tahun `ZScore_7-12` `ZScore_13-15` `ZScore_16-18` `ZScore_19-24`
##          <int>         <int>          <int>          <int>          <int>
## 1            0             0              0              0              0
```

3. Cleaning Data for Relationship Between Educational Attainment and Poverty in Indonesia in 2022

```
library(dplyr)
data = read.table(file.choose(), sep=",", header = T, skip = 1)

# looking for missing data
missing_summary <- data %>%
  summarise(across(everything(), ~ sum(is.na(.))))
print(missing_summary)
```

```
##   ACEH X99.44 X99.45 X99.08 X93.43 X97.63 X94.55 X74.36 X70.67 X74.46 X89.25
## 1    0      0      0      0      0      0      0      0      0      0      0
```

```
summary(data)
```

```
##       ACEH               X99.44           X99.45           X99.08           X93.43
##  Length:33          Min.   :78.43    Min.   :81.99    Min.   :80.09    Min.   :66.06
##  Class :character   1st Qu.:95.77    1st Qu.:96.94    1st Qu.:96.01    1st Qu.:86.09
##  Mode  :character   Median :97.02    Median :97.65    Median :97.76    Median :89.49
##                     Mean   :96.27    Mean   :96.92    Mean   :96.68    Mean   :88.17
##                     3rd Qu.:98.33    3rd Qu.:98.47    3rd Qu.:98.42    3rd Qu.:91.35
##                     Max.   :99.26    Max.   :99.18    Max.   :99.09    Max.   :95.34
##       X97.63           X94.55           X74.36           X70.67           X74.46           X89.25
##  Min.   :66.16    Min.   :67.12    Min.   :32.95    Min.   :38.47    Min.   :39.50    Min.   :62.39
##  1st Qu.:87.79    1st Qu.:88.08    1st Qu.:61.04    1st Qu.:58.75    1st Qu.:59.99    1st Qu.:82.26
##  Median :90.47    Median :90.05    Median :65.71    Median :66.02    Median :66.62    Median :84.45
##  Mean   :89.32    Mean   :89.40    Mean   :65.24    Mean   :64.50    Mean   :65.55    Mean   :83.58
##  3rd Qu.:92.84    3rd Qu.:92.95    3rd Qu.:69.43    3rd Qu.:67.81    3rd Qu.:68.96    3rd Qu.:85.64
##  Max.   :97.06    Max.   :97.02    Max.   :90.12    Max.   :87.92    Max.   :89.69    Max.   :94.63
```

```
column_types <- sapply(data, class)

# replace data with "Unknown" if missing data is char
dataNew <- data %>%
  mutate_if(is.character, ~ifelse(is.na(.), "Unknown", .))

# replace data with column's median if missing data is numerical
dataNew <- data %>%
  mutate_all(
    funs(if(is.numeric(.)) replace(., is.na(.), median(., na.rm = TRUE)) else .)
  )
```

```
## Warning: `funs()` was deprecated in dplyr 0.8.0.
## ℹ Please use a list of either functions or lambdas:
##
## # Simple named list: list(mean = mean, median = median)
##
## # Auto named with `tibble::lst()`: tibble::lst(mean, median)
##
## # Using lambdas list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.
```

```
# Looking for Duplicated Data
duplicated_rows <- duplicated(data)
duplicated_rows
```

```
##  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [16] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [31] FALSE FALSE FALSE
```

```
# Make a new data without duplicates in variabel called unique data specific
unique_data_specific <- distinct(data, .keep_all = TRUE)
unique_data_specific
```

```
##                    ACEH X99.44 X99.45 X99.08 X93.43 X97.63 X94.55 X74.36 X70.67 X74.46      X89.25
## 1        SUMATERA UTARA  98.57  98.74  98.75  91.35  92.84  94.35  72.81  77.16  74.43 89.58000
## 2        SUMATERA BARAT  95.29  97.87  95.81  89.49  88.83  90.65  70.06  65.96  68.64 84.22000
```

```
##  3                  RIAU 96.91 98.20 98.09 87.11 88.53 90.52 68.94 66.91 67.79 84.54667
##  4                 JAMBI 98.54 97.62 97.76 89.00 86.57 89.35 64.51 65.85 66.62 83.34667
##  5       SUMATERA SELATAN 97.82 97.53 97.58 87.68 88.41 87.95 67.20 67.07 64.81 84.33667
##  6              BENGKULU 98.16 97.65 97.10 89.94 90.81 89.25 62.46 64.88 63.41 84.44667
##  7              LAMPUNG 98.33 98.41 98.67 89.46 90.99 87.67 60.09 62.42 64.54 83.94000
##  8  KEP. BANGKA BELITUNG 96.61 96.45 96.01 80.99 84.72 87.11 63.98 66.87 68.96 82.68000
##  9              KEP. RIAU 98.16 98.38 97.92 92.71 95.72 95.51 81.07 73.93 78.97 89.34333
## 10           DKI JAKARTA 99.26 98.58 98.66 95.00 95.40 95.85 84.98 87.71 88.10 93.89667
## 11           JAWA BARAT 98.45 99.08 99.09 88.18 89.29 91.42 64.89 67.05 66.47 85.14000
## 12           JAWA TENGAH 98.06 98.01 98.42 88.44 90.02 90.64 59.90 58.75 58.35 82.26000
## 13        DI YOGYAKARTA 98.48 98.91 98.95 94.94 97.06 97.02 90.12 87.92 89.69 94.63000
## 14           JAWA TIMUR 97.76 98.71 98.78 90.30 90.47 90.74 66.33 66.87 68.65 85.35000
## 15              BANTEN 98.82 97.24 97.15 90.63 92.65 90.86 66.90 66.02 70.07 85.30333
## 16                BALI 97.02 97.55 98.43 94.26 94.15 93.03 75.86 76.59 76.51 89.43000
## 17  NUSA TENGGARA BARAT 98.71 98.47 98.11 92.19 95.39 92.95 65.71 61.00 63.66 84.95333
## 18  NUSA TENGGARA TIMUR 91.84 92.35 93.41 78.83 83.25 82.48 44.88 38.47 43.46 71.35667
## 19    KALIMANTAN BARAT 94.29 95.13 95.33 79.65 81.82 81.56 54.27 58.40 55.58 78.45000
## 20    KALIMANTAN TENGAH 97.45 98.51 97.47 89.76 87.79 88.92 61.04 61.88 63.93 82.72667
## 21    KALIMANTAN SELATAN 95.67 94.68 95.99 84.06 87.95 88.19 63.59 67.81 68.35 83.48000
## 22    KALIMANTAN TIMUR 96.82 99.18 97.88 95.34 95.32 94.85 74.26 74.00 73.63 89.50000
## 23    KALIMANTAN UTARA 95.77 96.94 96.41 90.14 90.55 88.08 62.30 54.80 59.50 80.76333
## 24      SULAWESI UTARA 96.10 96.74 96.18 91.05 91.98 92.07 68.56 66.66 67.57 85.12667
## 25      SULAWESI TENGAH 96.19 97.19 97.56 85.42 88.90 90.05 61.16 53.73 55.69 79.94000
## 26      SULAWESI SELATAN 97.30 98.05 98.37 88.18 90.55 88.74 69.43 68.32 67.41 85.64000
## 27    SULAWESI TENGGARA 95.58 97.24 97.83 90.88 91.19 89.55 70.65 65.97 68.28 84.80000
## 28            GORONTALO 93.44 95.12 93.69 81.22 80.56 83.71 53.73 45.12 46.19 73.60000
## 29      SULAWESI BARAT 95.93 97.15 95.13 86.09 84.14 84.04 56.22 55.18 54.79 78.82333
## 30              MALUKU 98.50 98.98 98.69 93.08 93.65 93.90 68.12 72.08 75.01 88.23667
## 31        MALUKU UTARA 96.97 97.72 98.30 92.93 94.92 93.46 66.95 67.10 64.61 86.58000
## 32          PAPUA BARAT 91.81 93.94 92.69 85.18 87.03 88.63 59.08 57.07 59.99 79.34667
## 33                PAPUA 78.43 81.99 80.09 66.06 66.16 67.12 32.95 39.01 39.50 62.38667
```

```r
# Function to detect outliers using IQR method
outliers_iqr <- function(df) {
  df %>%
    summarise(across(where(is.numeric), ~{
      Q1 <- quantile(., 0.25, na.rm = TRUE)
      Q3 <- quantile(., 0.75, na.rm = TRUE)
      IQR <- Q3 - Q1
      lower_bound <- Q1 - 1.5 * IQR
      upper_bound <- Q3 + 1.5 * IQR
      sum(. < lower_bound | . > upper_bound, na.rm = TRUE)
    }, .names = "IQR_{.col}"))
}

# Function to detect outliers using Z-Score method
outliers_zscore <- function(df, threshold = 3) {
  df %>%
    summarise(across(where(is.numeric), ~{
      z_scores <- scale(.)
      sum(abs(z_scores) > threshold, na.rm = TRUE)
    }, .names = "ZScore_{.col}"))
}
```

```r
# Detecting outliers
iqr_outliers_summary <- outliers_iqr(data)
zscore_outliers_summary <- outliers_zscore(data)

# Print
print(iqr_outliers_summary)
```

```
##   IQR_X99.44 IQR_X99.45 IQR_X99.08 IQR_X93.43 IQR_X97.63 IQR_X94.55 IQR_X74.36 IQR_X70.67
## 1          3          3          1          1          1          1          4          5
##   IQR_X74.46 IQR_X89.25
## 1          5          5
```

```r
print(zscore_outliers_summary)
```

```
##   ZScore_X99.44 ZScore_X99.45 ZScore_X99.08 ZScore_X93.43 ZScore_X97.63 ZScore_X94.55
## 1             1             1             1             1             1             1
##   ZScore_X74.36 ZScore_X70.67 ZScore_X74.46 ZScore_X89.25
## 1             1             0             0             1
```

4. Cleaning Data for Distribution of Libraries in Indonesia for the Academic Year 2022/2023

```r
library(dplyr)
data = read.table(file.choose(), sep=",", header = T, skip = 1)

# looking for missing data
missing_summary <- data %>%
  summarise(across(everything(), ~ sum(is.na(.))))
print(missing_summary)
```

```
##   Prov..D.K.I..Jakarta X1831 X264 X70 X45 X0 X2210 X950 X119 X25 X7 X0.1 X1101 X435 X45.1 X7.1 X4
## 1                    0     4    4    4    4  4     4    4    4    4  4    4     4    4     4    4  4
##   X0.2 X491 X469 X91 X11 X6 X0.3 X577 X4379
## 1    4    4    4   4   4  4    4    4     4
```

```r
summary(data)
```

```
##   Prov..D.K.I..Jakarta      X1831           X264           X70            X45             X0
##   Length:37            Min.   : 187   Min.   :  93   Min.   :  39   Min.   :  40.0   Min.   :0
##   Class :character     1st Qu.: 529   1st Qu.: 283   1st Qu.: 230   1st Qu.: 146.0   1st Qu.:0
##   Mode  :character     Median : 995   Median : 590   Median : 381   Median : 208.0   Median :0
##                        Mean   :1463   Mean   : 936   Mean   : 590   Mean   : 347.5   Mean   :0
##                        3rd Qu.:1573   3rd Qu.: 982   3rd Qu.: 655   3rd Qu.: 340.0   3rd Qu.:0
##                        Max.   :6550   Max.   :4705   Max.   :2525   Max.   :1447.0   Max.   :0
##                        NA's   :4      NA's   :4      NA's   :4      NA's   :4        NA's   :4
##       X2210            X950            X119            X25             X7             X0.1
##   Min.   :  359   Min.   :  95.0   Min.   :  30.0   Min.   :  13.0   Min.   :  12.00   Min.   :0
##   1st Qu.: 1265   1st Qu.: 262.0   1st Qu.:  98.0   1st Qu.:  68.0   1st Qu.:  39.00   1st Qu.:0
```

```
##   Median : 2174   Median : 358.0   Median : 179.0   Median :119.0   Median : 74.00   Median :0
##   Mean   : 3336   Mean   : 576.5   Mean   : 278.7   Mean   :166.4   Mean   : 90.03   Mean   :0
##   3rd Qu.: 3378   3rd Qu.: 621.0   3rd Qu.: 315.0   3rd Qu.:202.0   3rd Qu.:112.00   3rd Qu.:0
##   Max.   :15022   Max.   :2692.0   Max.   :1306.0   Max.   :665.0   Max.   :313.00   Max.   :0
##   NA's   :4       NA's   :4        NA's   :4        NA's   :4       NA's   :4        NA's   :4
##        X1101            X435            X45.1            X7.1             X4              X0.2
##   Min.   : 160   Min.   : 31.0   Min.   : 4.00   Min.   : 2.00   Min.   : 2.00   Min.   :0
##   1st Qu.: 455   1st Qu.: 123.0   1st Qu.: 36.00   1st Qu.: 20.00   1st Qu.: 10.00   1st Qu.:0
##   Median : 712   Median : 144.0   Median : 60.00   Median : 32.00   Median : 19.00   Median :0
##   Mean   :1112   Mean   : 243.5   Mean   : 89.58   Mean   : 44.52   Mean   : 25.97   Mean   :0
##   3rd Qu.:1242   3rd Qu.: 315.0   3rd Qu.:111.00   3rd Qu.: 62.00   3rd Qu.: 37.00   3rd Qu.:0
##   Max.   :4961   Max.   :1011.0   Max.   :377.00   Max.   :148.00   Max.   :100.00   Max.   :0
##   NA's   :4      NA's   :4        NA's   :4        NA's   :4        NA's   :4        NA's   :4
##        X491            X469             X91              X11              X6              X0.3
##   Min.   :  54.0   Min.   :  14.0   Min.   :  5.00   Min.   :  3.00   Min.   :  0.00   Min.   :0
##   1st Qu.: 182.0   1st Qu.:  60.0   1st Qu.: 21.00   1st Qu.: 12.00   1st Qu.:  4.00   1st Qu.:0
##   Median : 240.0   Median : 107.0   Median : 43.00   Median : 17.00   Median :  9.00   Median :0
##   Mean   : 403.6   Mean   : 240.5   Mean   : 93.42   Mean   : 31.79   Mean   : 14.21   Mean   :0
##   3rd Qu.: 530.0   3rd Qu.: 174.0   3rd Qu.: 64.00   3rd Qu.: 34.00   3rd Qu.: 18.00   3rd Qu.:0
##   Max.   :1609.0   Max.   :1707.0   Max.   :695.00   Max.   :191.00   Max.   : 77.00   Max.   :0
##   NA's   :4        NA's   :4        NA's   :4        NA's   :4        NA's   :4        NA's   :4
##        X577            X4379
##   Min.   :  22.0   Min.   :  595
##   1st Qu.: 105.0   1st Qu.: 1994
##   Median : 170.0   Median : 3211
##   Mean   : 379.9   Mean   : 5231
##   3rd Qu.: 271.0   3rd Qu.: 5768
##   Max.   :2670.0   Max.   :21762
##   NA's   :4        NA's   :4
```

```
column_types <- sapply(data, class)

# replace data with "Unknown" if missing data is char
dataNew <- data %>%
  mutate_if(is.character, ~ifelse(is.na(.), "Unknown", .))

# replace data with column's median if missing data is numerical
dataNew <- data %>%
  mutate_all(
    funs(if(is.numeric(.)) replace(., is.na(.), median(., na.rm = TRUE)) else .)
  )
```

```
## Warning: 'funs()' was deprecated in dplyr 0.8.0.
## i Please use a list of either functions or lambdas:
##
## # Simple named list: list(mean = mean, median = median)
##
## # Auto named with 'tibble::lst()': tibble::lst(mean, median)
##
## # Using lambdas list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was generated.
```

```r
# Looking for Duplicated Data
duplicated_rows <- duplicated(data)
duplicated_rows
```

```
##   [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [16] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [31] FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE
```

```r
# Make a new data without duplicates in variabel called unique data specific
unique_data_specific <- distinct(data, .keep_all = TRUE)
unique_data_specific
```

```
##                 Prov..D.K.I..Jakarta X1831 X264 X70 X45 X0 X2210 X950 X119 X25 X7 X0.1 X1101 X435
## 1                  Prov. Jawa Barat  4890 3673 2525 1434  0 12522 2692 1306 665 298    0  4961 1011
## 2                 Prov. Jawa Tengah  6550 4705 2436 1331  0 15022 1780  934 489 205    0  3408  590
## 3              Prov. D.I. Yogyakarta   995  487  197   83  0  1762  282   98  43  42    0   465  144
## 4                  Prov. Jawa Timur  6345 3985 2054 1447  0 13831 2304 1201 559 313    0  4377  909
## 5                       Prov. Aceh  1275  847  642  340  0  3104  539  307 230 114    0  1190  272
## 6               Prov. Sumatera Utara  3348 1945 1028  618  0  6939 1319  568 328 194    0  2409  673
## 7               Prov. Sumatera Barat  1573  976  526  250  0  3325  485  191 119  61    0   856  216
## 8                     Prov. Riau  1272  666  392  207  0  2537  613  209 135  91    0  1048  316
## 9                     Prov. Jambi   783  590  376  195  0  1944  262  179 115  74    0   630  137
## 10            Prov. Sumatera Selatan  1822 1068  655  315  0  3860  627  315 204  96    0  1242  351
## 11                  Prov. Lampung  1390  982  734  418  0  3524  621  347 202 112    0  1282  274
## 12            Prov. Kalimantan Barat  1279 1014  685  400  0  3378  584  333 185 112    0  1214  218
## 13           Prov. Kalimantan Tengah   813  556  316  201  0  1886  358  178  98  72    0   706  141
## 14          Prov. Kalimantan Selatan  1068  639  381  200  0  2288  357  153  93  34    0   637  149
## 15            Prov. Kalimantan Timur   833  345  155  146  0  1479  381  132  67  39    0   619  134
## 16              Prov. Sulawesi Utara   731  411  289  241  0  1672  298  181 144  89    0   712  136
## 17             Prov. Sulawesi Tengah   999  547  437  285  0  2268  376  185 128 105    0   794  157
## 18             Prov. Sulawesi Selatan  2135 1587 1181  618  0  5521  748  375 319 145    0  1587  333
## 19           Prov. Sulawesi Tenggara   850  444  414  280  0  1988  379  166 137  88    0   770  145
## 20                    Prov. Maluku   568  245  311  208  0  1332  265  102  78  75    0   520  132
## 21                     Prov. Bali  1260  603  230   81  0  2174  286   75  31  17    0   409  125
## 22         Prov. Nusa Tenggara Barat   899  765  621  251  0  2536  340  230 166  52    0   788  132
## 23         Prov. Nusa Tenggara Timur  1627 1189  895  596  0  4307  823  400 254 149    0  1626  315
## 24                    Prov. Papua   434  283  219  157  0  1093  225  169  88  68    0   550   89
## 25                  Prov. Bengkulu   506  359  307   93  0  1265  199  137  85  34    0   455   84
## 26               Prov. Maluku Utara   284  221  283  147  0   935  142   95 109  51    0   397   95
## 27                    Prov. Banten  1661  793  399  337  0  3190  797  291 202  78    0  1368  396
## 28 Prov.  Kepulauan Bangka Belitung   400  183  138  114  0   835  150   44  13  13    0   220   45
## 29                  Prov. Gorontalo   414  189  192   79  0   874  183   65  52  30    0   330   54
## 30              Prov. Kepulauan Riau   529  145   77   78  0   829  263   58  22  12    0   355  123
## 31                Prov. Papua Barat   237  136   89   76  0   538  124   68  44  34    0   270   58
## 32              Prov. Sulawesi Barat   313  216  246  200  0   975  126   76  68  58    0   328   52
## 33            Prov. Kalimantan Utara   187   93   39   40  0   359   95   30  19  16    0   160   31
## 34                                      NA   NA   NA   NA NA    NA   NA   NA  NA  NA   NA    NA   NA
##    X45.1 X7.1  X4 X0.2 X491 X469 X91 X11 X6 X0.3 X577 X4379
## 1    377  148  73    0 1609 1707 695 191 77    0 2670 21762
## 2    202   73  41    0  906 1039 400  99 41    0 1579 20915
## 3     30   22   8    0  204  163  51  12  6    0  232  2663
## 4    320  146 100    0 1475 1280 483 139 61    0 1963 21646
## 5    115   89  54    0  530  110  51  34 18    0  213  5037
```

```
## 6    238   75   57     0 1043   579  195   75 39     0   888 11279
## 7     61   38   31     0  346   146   41   17 10     0   214  4741
## 8     87   37   18     0  458   174   52   15  2     0   243  4286
## 9     60   28   14     0  239   101   43   21  5     0   170  2983
## 10   129   72   31     0  583   185   64   24  8     0   281  5966
## 11   111   86   42     0  513   270  117   43 19     0   449  5768
## 12    92   55   37     0  402   114   57   16  9     0   196  5190
## 13    60   20   13     0  234    70   30    9  6     0   115  2941
## 14    27   19    9     0  204    88   19    4  4     0   115  3244
## 15    59   23   10     0  226   107   62   14 12     0   195  2519
## 16    53   32   19     0  240   100   39   20 10     0   169  2793
## 17    43   24   13     0  237    86   28   27  8     0   149  3448
## 18   150   68   45     0  596   218  102   43 11     0   374  8078
## 19    63   59   47     0  314    73   22   25 19     0   139  3211
## 20    56   55   32     0  275    60   21   13  8     0   102  2229
## 21    25    4    6     0  160   129   32    6  1     0   168  2911
## 22    76   62   22     0  292   147   73   36 11     0   267  3883
## 23   142   57   32     0  546   166   51   25 29     0   271  6750
## 24    86   27   15     0  217    53   45   22 14     0   134  1994
## 25    36   23    7     0  150    53   27   16  4     0   100  1970
## 26    42   24   21     0  182    48   20   14  3     0    85  1599
## 27   110   54   20     0  580   418  192   53 18     0   681  5819
## 28     4   10    6     0   65    44    6    6  0     0    56  1176
## 29    11    4    2     0   71    37   13    3  2     0    55  1330
## 30    18    8    6     0  155    80   19    5  1     0   105  1444
## 31    40   15   10     0  123    23   13    4  4     0    44   975
## 32    18   10   10     0   90    53   15   15  9     0    92  1485
## 33    15    2    6     0   54    14    5    3  0     0    22   595
## 34    NA   NA   NA    NA   NA    NA   NA   NA NA    NA    NA    NA
```

```r
# Function to detect outliers using IQR method
outliers_iqr <- function(df) {
  df %>%
    summarise(across(where(is.numeric), ~{
      Q1 <- quantile(., 0.25, na.rm = TRUE)
      Q3 <- quantile(., 0.75, na.rm = TRUE)
      IQR <- Q3 - Q1
      lower_bound <- Q1 - 1.5 * IQR
      upper_bound <- Q3 + 1.5 * IQR
      sum(. < lower_bound | . > upper_bound, na.rm = TRUE)
    }, .names = "IQR_{.col}"))
}

# Function to detect outliers using Z-Score method
outliers_zscore <- function(df, threshold = 3) {
  df %>%
    summarise(across(where(is.numeric), ~{
      z_scores <- scale(.)
      sum(abs(z_scores) > threshold, na.rm = TRUE)
    }, .names = "ZScore_{.col}"))
}

# Detecting outliers
iqr_outliers_summary <- outliers_iqr(data)
```

```r
zscore_outliers_summary <- outliers_zscore(data)

# Print
print(iqr_outliers_summary)
```

```
##   IQR_X1831 IQR_X264 IQR_X70 IQR_X45 IQR_X0 IQR_X2210 IQR_X950 IQR_X119 IQR_X25 IQR_X7 IQR_X0.1
## 1         4        3       3       3      0         4        4        3       3      2         0
##   IQR_X1101 IQR_X435 IQR_X45.1 IQR_X7.1 IQR_X4 IQR_X0.2 IQR_X491 IQR_X469 IQR_X91 IQR_X11 IQR_X6
## 1         3        3         3        2      1        0        2        5       5       4      3
##   IQR_X0.3 IQR_X577 IQR_X4379
## 1        0        5         3
```

```r
print(zscore_outliers_summary)
```

```
##   ZScore_X1831 ZScore_X264 ZScore_X70 ZScore_X45 ZScore_X0 ZScore_X2210 ZScore_X950 ZScore_X119
## 1            2           1          1          0         0            1           1           1
##   ZScore_X25 ZScore_X7 ZScore_X0.1 ZScore_X1101 ZScore_X435 ZScore_X45.1 ZScore_X7.1 ZScore_X4
## 1          1         1           0            1           1           1           0          1
##   ZScore_X0.2 ZScore_X491 ZScore_X469 ZScore_X91 ZScore_X11 ZScore_X6 ZScore_X0.3 ZScore_X577
## 1           0           1           1          1          1         1           0           1
##   ZScore_X4379
## 1            0
```

================ Visualization using Shiny ================

5. Visualization using Shiny

```r
library(shiny)
library(bslib)
library(ggplot2)
library(plotly)
library(dplyr)
library(readxl)


data = read.table(file.choose(), sep=",", header = T)
data2 = read.table(file.choose(), sep=",", header = T)
data3 = read.table(file.choose(), sep=",", header = T)
data4 = read.table(file.choose(), sep=",", header = T)

dataPartisipasi = read_excel(file.choose())

education_2022 = read.table(file.choose(), sep=",", header = T)
poverty_2022 = read.table(file.choose(), sep=",", header = T)

education_2022 <- na.omit(education_2022)
poverty_2022 <- na.omit(poverty_2022)

education_2022$Provinsi <- tolower(education_2022$Provinsi)
poverty_2022$Provinsi <- tolower(poverty_2022$Provinsi)
```

```
education_2022$Provinsi[education_2022$Provinsi == "kep. bangka belitung"] <- "kepulauan bangka belitun
education_2022$Provinsi[education_2022$Provinsi == "kep. riau"] <- "kepulauan riau"
education_2022$Provinsi[education_2022$Provinsi == "di yogyakarta"] <- "di. yogyakarta"
final <- merge(poverty_2022,education_2022, by.x= "Provinsi")


library_data <- read.table(file.choose(), sep=",", header = T)

library_data <- library_data %>%
  select(Province = `Provinsi`, Total_perpustakaan_per_provinsi = `Total.perpustakaan.per.provinsi`) %>
  filter(!is.na(Total_perpustakaan_per_provinsi))    %>%
  mutate(
    Province = gsub("Prov. ", "", Province),
    Total_perpustakaan_per_provinsi = as.numeric(Total_perpustakaan_per_provinsi)
  )

ui <- navbarPage(
  title = "Quality of Education in Indonesia in 2022",
  theme = bs_theme(bootswatch = "morph"),
  navbarMenu(
    title = "Plots",
    tabPanel(
      title = "Distribution of Teachers in Indonesia in 2022",
      fluidPage(
        fluidRow(
          titlePanel("Distribution of Teachers in Indonesia in 2022"),
          column(
            width = 6,
            selectInput(
              inputId = "x",
              label = "Choice of Option:",
              choices = unique(data$Provinsi),
              selected = "Aceh"
            )
          )
        ),
        fluidRow(
          column(
            width = 12,
            plotlyOutput(outputId = "barplot")
          )
        )
      )
    ),

    tabPanel(
      title = "School Participation Rates in Indonesia from 2002 to 2022",
      fluidPage(
        fluidRow(
          titlePanel("School Participation Rates in Indonesia from 2002 to 2022")
        ),
        fluidRow(
          column(
```

```
            width = 12,
            plotlyOutput(outputId = "lineplot")
          )
        )
      )
    ),

    tabPanel(
      title = "Relationship Between Educational Attainment and Poverty in Indonesia in 2022",
      fluidPage(
        fluidRow(
          titlePanel("Relationship Between Educational Attainment and Poverty in Indonesia in 2022"),
        ),
        fluidRow(
          column(
            width = 12,
            plotlyOutput(outputId = "scatter")
          )
        )
      )
    ),

    tabPanel(
      title = "Distribution of Libraries in Indonesia for the Academic Year 2022/2023",
      fluidPage(
        fluidRow(
          titlePanel("Distribution of Libraries in Indonesia for the Academic Year 2022/2023"),
          column(
            width = 6,
            selectInput(
              inputId = "library_choice",
              label = "Choose an Option:",
              choices = c("5 Highest", "5 Lowest"),
              selected = "5 Highest"
            )
          )
        ),
        fluidRow(
          column(
            width = 12,
            plotlyOutput(outputId = "library_plot")
          )
        )
      )
    )

  )
)

server <- function(input, output) {
  output$barplot <- renderPlotly({
    plot_ly(
      data = subset(data, Provinsi %in% input$x),
```

```
      y = data[data$Provinsi %in% input$x, "Jumlah.Guru.SD..Negeri.Swasta."],
      x = subset(data, Provinsi %in% input$x)$Provinsi,
      type = "bar",
      color = "blue",
      name = "Total SD Teachers"
    ) %>%
      add_bars(
        data = subset(data2, Provinsi %in% input$x),
        x = subset(data2, Provinsi %in% input$x)$Provinsi,
        y = data2[data2$Provinsi %in% input$x, "Jumlah.Guru.SMP..Negeri.Swasta."],
        color = "orange",
        name = "Total SMP Teachers"
      ) %>%
      add_bars(
        data = subset(data3, Provinsi %in% input$x),
        x = subset(data3, Provinsi %in% input$x)$Provinsi,
        y = data3[data3$Provinsi %in% input$x, "Jumlah.Guru.SMA..Negeri.Swasta."],
        color = "red",
        name = "Total SMA Teachers"
      ) %>%
      add_bars(
        data = subset(data4, Provinsi %in% input$x),
        x = subset(data4, Provinsi %in% input$x)$Provinsi,
        y = data4[data4$Provinsi %in% input$x, "Jumlah.Guru.SMK..Negeri.Swasta."],
        color = "green",
        name = "Total SMK Teachers"
      )
})

output$lineplot <- renderPlotly({
  plot_ly(
    data = dataPartisipasi,
    x = ~dataPartisipasi$Tahun,
    y = ~dataPartisipasi$`7-12`,
    type = 'scatter',
    mode = 'lines',
    name = '7-12'
  ) %>%
    add_trace(
      y = ~dataPartisipasi$`13-15`,
      name = '13-15',
      mode = 'lines'
    ) %>%
    add_trace(
      y = ~dataPartisipasi$`16-18`,
      name = '16-18',
      mode = 'lines'
    ) %>%
    add_trace(
      y = ~dataPartisipasi$`19-24`,
      name = '19-24',
      mode = 'lines'
    ) %>%
```

```
      layout(
        title = "Participation by Age Group",
        xaxis = list(title = "Year"),
        yaxis = list(title = "Participation Percentage")
      )
  })

 output$scatter <- renderPlotly({
  lm_model <- lm(Rata.rata.kemiskinan.tahun.2022 ~ Rata.rata.penyelesaian.pendidikan.tahun.2022, data =
  final$trendline <- predict(lm_model)

  cor_test <- cor.test(final$Rata.rata.penyelesaian.pendidikan.tahun.2022, final$Rata.rata.kemiskinan.t
  cor_value <- round(cor_test$estimate, 2)
  p_value <- format.pval(cor_test$p.value, digits = 3)

  plot_ly(
    data = final,
    x = ~Rata.rata.penyelesaian.pendidikan.tahun.2022,
    y = ~Rata.rata.kemiskinan.tahun.2022,
    type = 'scatter',
    mode = 'markers',
    text = ~Provinsi,
    name = 'Correlation',
    marker = list(size = 10)
  ) %>%
    add_lines(
      x = ~Rata.rata.penyelesaian.pendidikan.tahun.2022,
      y = ~trendline,
      line = list(color = 'darkorange'),
      name = 'Trendline'
    ) %>%
    layout(
      title = paste("Relationship Between Educational Attainment and Poverty in Indonesia in 2022",
                    "<br>Correlation: ", cor_value, ", p-value: ", p_value),
      xaxis = list(title = "Average Educational Attainment (2022)"),
      yaxis = list(title = "Average Poverty Rate (2022)")
    )
})

 output$library_plot <- renderPlotly({
   library_data_sorted <- library_data %>%
     arrange(desc(Total_perpustakaan_per_provinsi))

   top5_highest <- library_data_sorted %>%
     head(5)

   top5_lowest <- library_data_sorted %>%
     tail(5)

   selected_data <- if(input$library_choice == "5 Highest"){
     top5_highest
   }else {
     top5_lowest
```

```
  }

  plot_ly(
    data = selected_data,
    x = ~Province,
    y = ~Total_perpustakaan_per_provinsi,
    type = 'bar',
    color = ~Province,
    text = ~Total_perpustakaan_per_provinsi
  ) %>%
    layout(
      title = paste(input$library_choice, "Distribution of Libraries in Indonesia for the  Academic Yea
      xaxis = list(title = "Province"),
      yaxis = list(title = "Total Libraries")
    )
 })

}
shinyApp(ui = ui, server = server)
```

Data Preparation

The app starts by loading several datasets using the read.table and read_excel functions. These datasets include information about the distribution of teachers and libraries, school participation rates, educational attainment, and poverty rates in various provinces of Indonesia. Data cleaning steps are performed, such as removing missing values with na.omit, converting province names to lowercase for consistency, and merging the education and poverty datasets on the province column.

Data Transformation

For the library data, the app selects and filters relevant columns, removes any "Prov." prefixes from province names, and abbreviates "Kepulauan" to "Kep." using the gsub function. The numeric values for the total number of libraries are also converted to ensure proper plotting.

User Interface

The UI is created using navbarPage from the Shiny library, with a bootswatch theme applied using bs_theme. The UI is divided into multiple tabs, each corresponding to a different plot:

- Distribution of Teachers: Users can select a province to view the distribution of teachers at various school levels (SD, SMP, SMA, SMK) using a bar plot.
- School Participation Rates: A line plot shows school participation rates for different age groups over the years 2002 to 2022.
- Educational Attainment and Poverty: A scatter plot displays the relationship between educational attainment and poverty rates, with a trendline and correlation information.
- Distribution of Libraries: Users can choose to see the top 5 provinces with the highest or lowest number of libraries in a bar plot.

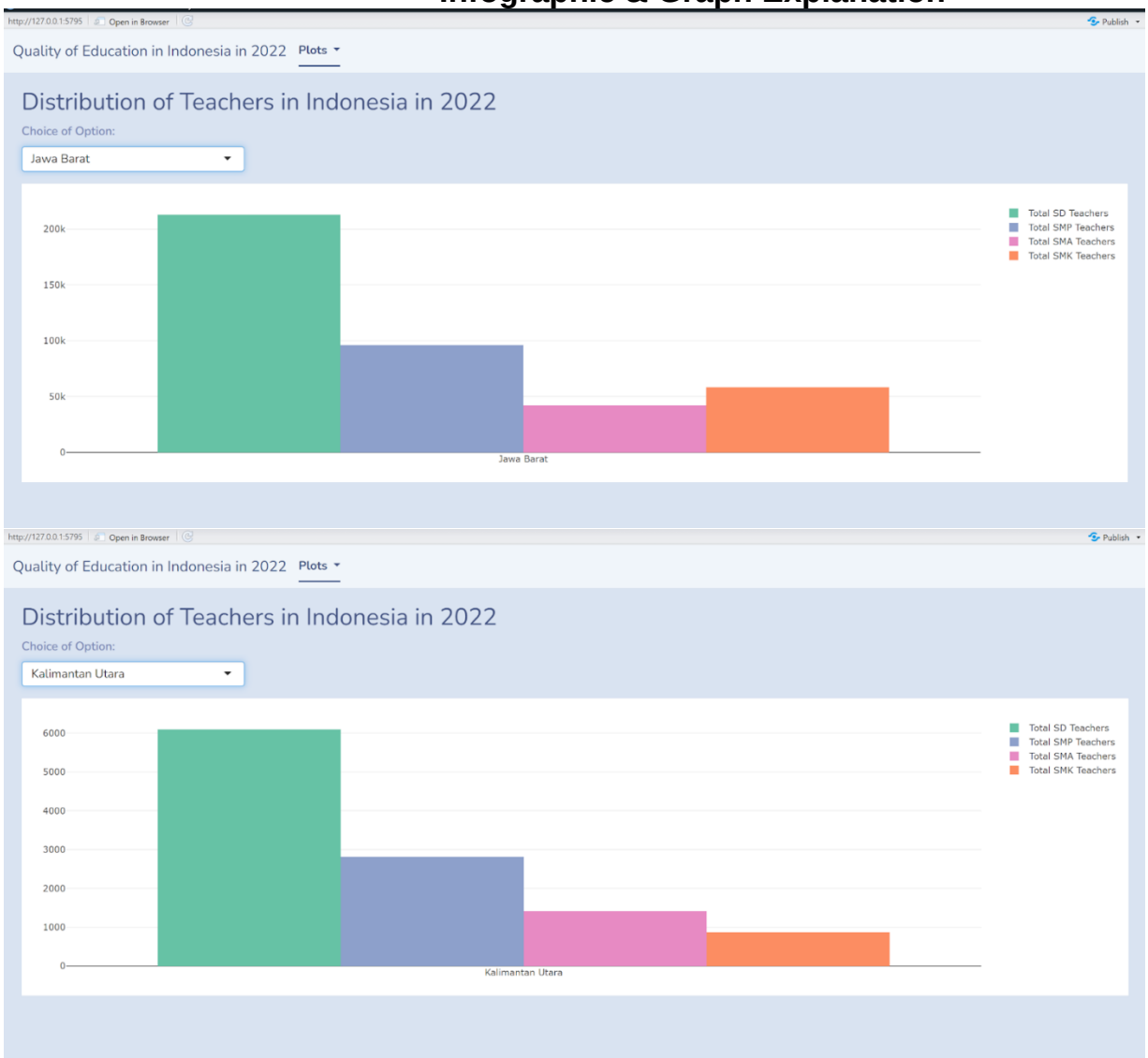
Server Logic

The server logic handles the rendering of the plots based on user inputs:

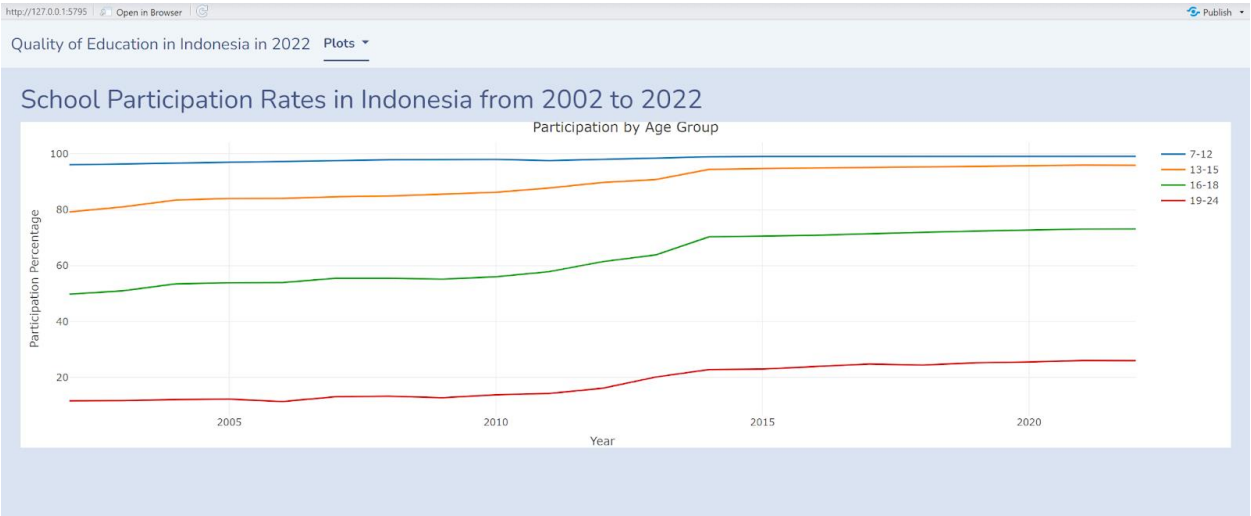
- Bar Plot for Teachers: Uses plot_ly to create bar plots showing the number of teachers at different school levels. The plot updates based on the selected province.

- Line Plot for Participation Rates: Plots participation rates for various age groups over time using plot_ly.
- Scatter Plot for Educational Attainment and Poverty: A scatter plot with a trendline is created using plot_ly, showing the relationship between educational attainment and poverty rates. The correlation coefficient and p-value are displayed in the plot title.
- Bar Plot for Libraries: Based on the user's choice of top 5 highest or lowest, the app sorts the data and creates a bar plot using plot_ly.

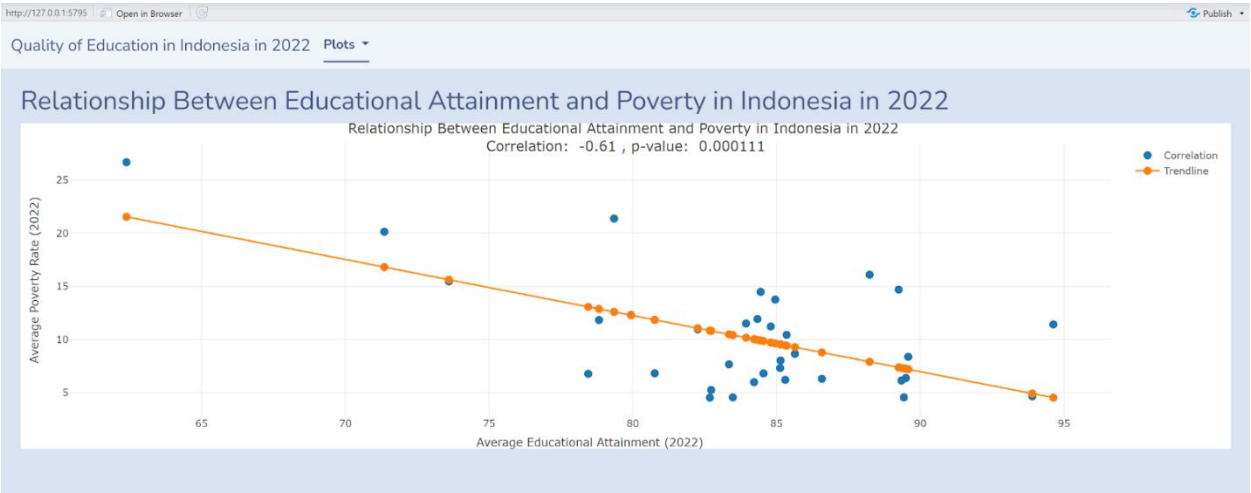
========== Infographic & Graph Explanation==========



The first graph discusses the distribution of teachers in each province of Indonesia in 2022. We use a bar chart to make it easier to see the distribution of teachers. Based on this graph, we find that the highest teacher distribution is in West Java province, and the lowest teacher distribution is in North Kalimantan province.



The second graph explains the growth in the percentage of school participation, showing the number of people still in school from 2002 to 2022. We use a line chart to make it easier to see the growth percentage. Based on this graph, we find that there is an increase in the percentage of people still in school, both male and female, across all age ranges from 7 to 24 years. This proves that there is growing awareness among the community about the importance of education.



The third graph discusses the relationship between educational attainment and poverty in Indonesia in 2022. We use a scatter plot to determine whether there is a relationship between educational attainment and poverty. Through this graph, we get a p-value of 0.000111. This value is smaller than the alpha value of 0.05, indicating that there is a correlation between poverty and the quality of education received.

The fourth graph discusses the distribution of libraries in Indonesia for the academic year 2022/2023. We use a bar chart to make it easier to see the distribution of libraries. Based on this graph, we can see the five regions with the most library distribution and the five regions with the least library distribution. From these two graphs, we find that West Java has the highest distribution of libraries with 21,762 libraries, and North Kalimantan has the lowest distribution of libraries with 595 libraries.

Based on the analysis of the four presented graphics, it is evident that education plays a crucial role in driving Indonesia's social and economic progress. The significant increase in the percentage of school participation across various age ranges highlights the advancements in the educational sector and the growing awareness of the importance of education among the population.

However, Indonesia still faces challenges in achieving equitable distribution of educational resources. The disparity in the distribution of teachers and libraries, particularly between regions like West Java and North Kalimantan, indicates a need for more focused efforts to ensure educational equity across the nation. Moreover, the correlation between education levels and poverty underscores the critical need to address educational disparities to alleviate poverty and promote inclusive growth.

In conclusion, while Indonesia has made noteworthy strides in improving educational access and quality, continued efforts are essential to address regional inequalities and ensure that all citizens, regardless of their location, have access to quality education. Achieving this will be instrumental in

fulfilling the targets of SDG 4 and fostering a more equitable and prosperous future for Indonesia. Thank you for joining us in this exploration of the educational landscape in Indonesia. We appreciate your attention and support. With that, we bid you farewell.

Source

1. Distribution of Teachers in Indonesia in 2022
https://www.bps.go.id/id/statistics-table/3/VWtKTmFFbDZaSFJWWVhOYU16WmhaRzlCYlM5Wlp6MDkjMw==/jumlah-sekolah--guru--dan-murid-sekolah-dasar--sd--di-bawah-kementerian-pendidikan-dan-kebudayaan-menurut-provinsi--2017-2018.html

https://www.bps.go.id/id/statistics-table/3/ZHpkb1ZtcDNZV2RHTlUweVdFZ3JhVkl3Ym1ScVp6MDkjMw==/jumlah-sekolah--guru--dan-murid-sekolah-menengah-pertama--smp--di-bawah-kementerian-pendidikan-dan-kebudayaan-menurut-provinsi--2019-2020.html

https://www.bps.go.id/id/statistics-table/3/YTFsRmNubEhOWE5ZTUZsdWVHOHhMMFpPWm5VMFp6MDkjMw==/jumlah-sekolah--guru--dan-murid-sekolah-menengah-atas--sma--di-bawah-kementerian-pendidikan-dan-kebudayaan-menurut-provinsi--2016-2017.html

https://www.bps.go.id/id/statistics-table/3/TVU5MFYwMVlaMFJ4ZW5obWJGZHNVMjFpVUhoMlp6MDkjMw==/jumlah-sekolah--guru--dan-murid-sekolah-menengah-kejuruan--smk--di-bawah-kementerian-pendidikan-dan-kebudayaan-menurut-provinsi--2019-2020.html

2. School Participation Rates in Indonesia from 2002 to 2022
https://www.bps.go.id/id/statistics-table/1/MTUzMyMx/persentase-penduduk-usia-7-24-tahun-menurut-jenis-kelamin-kelompok-umur-dan-partisipasi-sekolah-2002-2023.html

3. Relationship Between Educational Attainment and Poverty in Indonesia in 2022
https://www.bps.go.id/id/statistics-table/2/MTk4MCMy/tingkat-penyelesaian-pendidikan-menurut-jenjang-pendidikan-dan-provinsi.html

https://jatim.bps.go.id/indicator/23/344/1/persentase-penduduk-miskin-menurut-provinsi-.html

4. Distribution of Libraries in Indonesia for the Academic Year 2022/2023
https://data.kemdikbud.go.id/dataset/detail/33/L0-000000/2022/SD-1#filter-section