Exploring ASER2016 Pakistan Data

Yusei Hara

2020/10/1

ASER Pakistan 2016

In this piece of paper, a set of data obtained from Annual Status of Education Report (ASER) is explored. The raw data was downloaded from the link here. https://palnetwork.org/aser-centre/

Preparation

```
library(tidyverse)
library(ggplot2)
```

Packages Used

```
school <- read.csv("aser/ASER2016GSchool.csv")
child <- read.csv("aser/ASER2016Child.csv")
```

```
RegionName <- c("2" =  "Panjab",
                "3" = "Sindh",
                "4" = "Balochistan",
                "5" = "Khyber Pakhtunkhwa",
                "6" = "Gilgit-Baltistan",
                "7" = "Azad Jammu and Kashmir",
                "8" = "Islamabad - ICT",
                "9" = "Federally Administrated Tribal Areas")
Gender <- c("0" = "Male",
            "-1" = "Female")
```

Data Installation

Exploration

```r
length(unique(child$CID))
```

Checking Samplesizes

```
## [1] 255196
```

The whole samplesize (the numebr of children) of this dataset is 255196.

```r
child %>%
  filter(DID == 266) %>%
  summarize(N_hunza = length(unique(CID)))
```

```
##   N_hunza
## 1    1641
```

The samplesize of Hunza alone is 1641.

Exploration in Hunza

```r
child %>%
  filter(DID == 266) %>%
  summarize(gender_proportion = mean(C002))
```

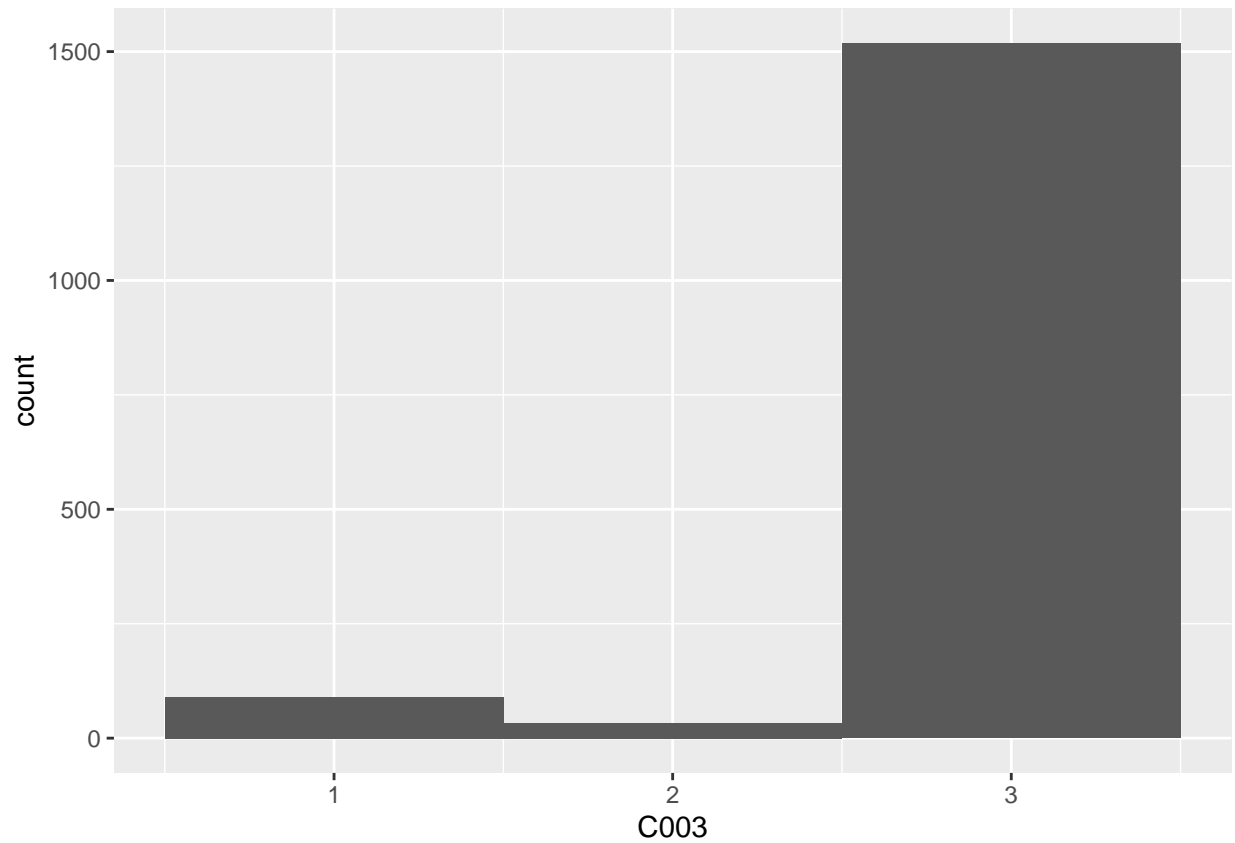**Gender Proportion**

```
##   gender_proportion
## 1        -0.5173675
```

-1: female, 0: male gender_proportion = -0.5173675 means there are a little more girls in the dataset.
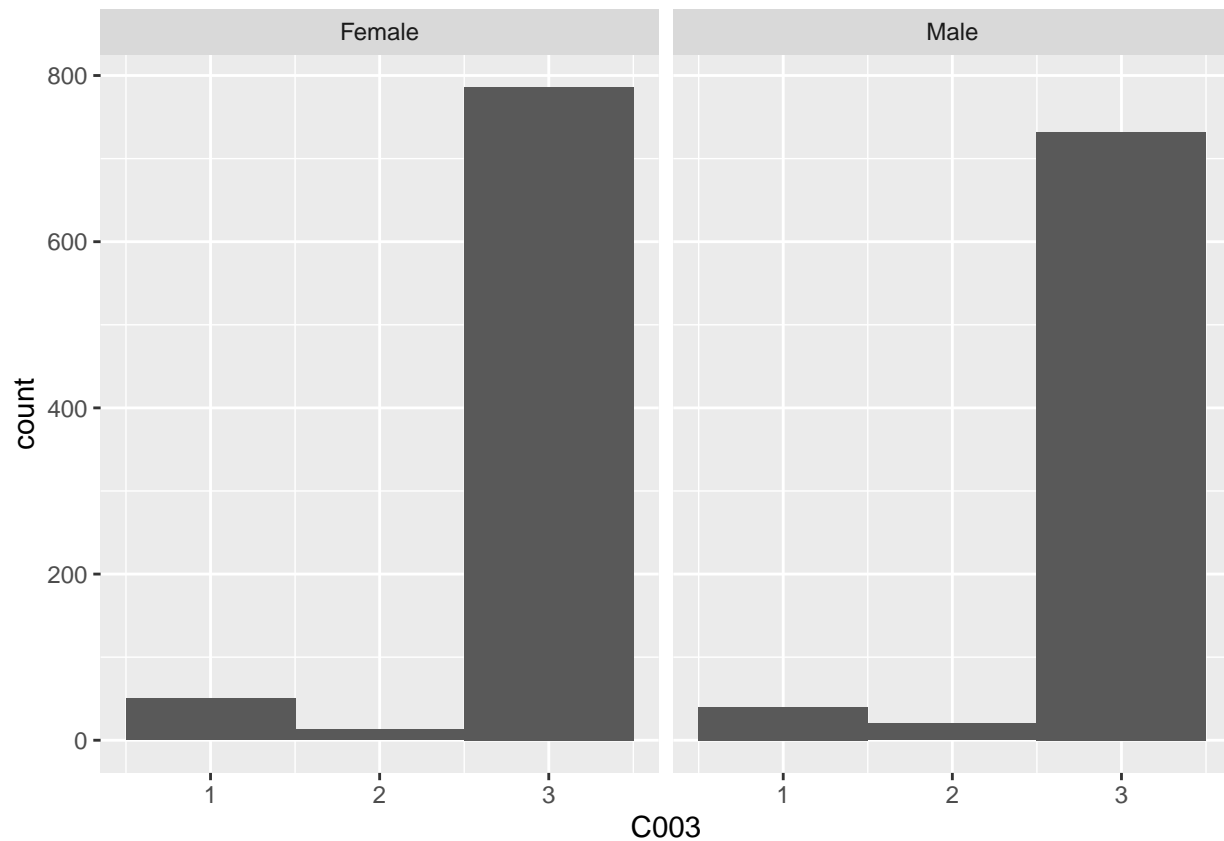
Eduation Status 1 = never enrolled; 2 = drop-out; 3 = currently enrolled

```r
child %>%
  filter(DID == 266) %>%
  ggplot(aes(C003)) +
  geom_histogram(bins = 3)
```

Education Status by Gender   Both genders look pretty good interms of the absolute number of currently-enrolled-children

```
child %>%
  filter(DID == 266) %>%
  ggplot(aes(C003)) +
  geom_histogram(bins = 3, binwidth = 1) +
  facet_grid(~C002, labeller = labeller(C002 = Gender))
```

The Enrollment Rate by Gender   As a rate, both are doing pretty good

```
child %>%
  filter(DID == 266) %>%
  group_by(C002) %>%
  summarize(enrollment_rate = mean(C003 == 3)) %>%
  ungroup() %>%
  ggplot(aes(C002, enrollment_rate)) +
  geom_col() +
  scale_y_continuous() +
  geom_label(aes(label = enrollment_rate)) +
  scale_x_continuous(breaks = c(-1, 0), labels = c("Female", "Male"))
```

## `summarise()` ungrouping output (override with `.groups` argument)
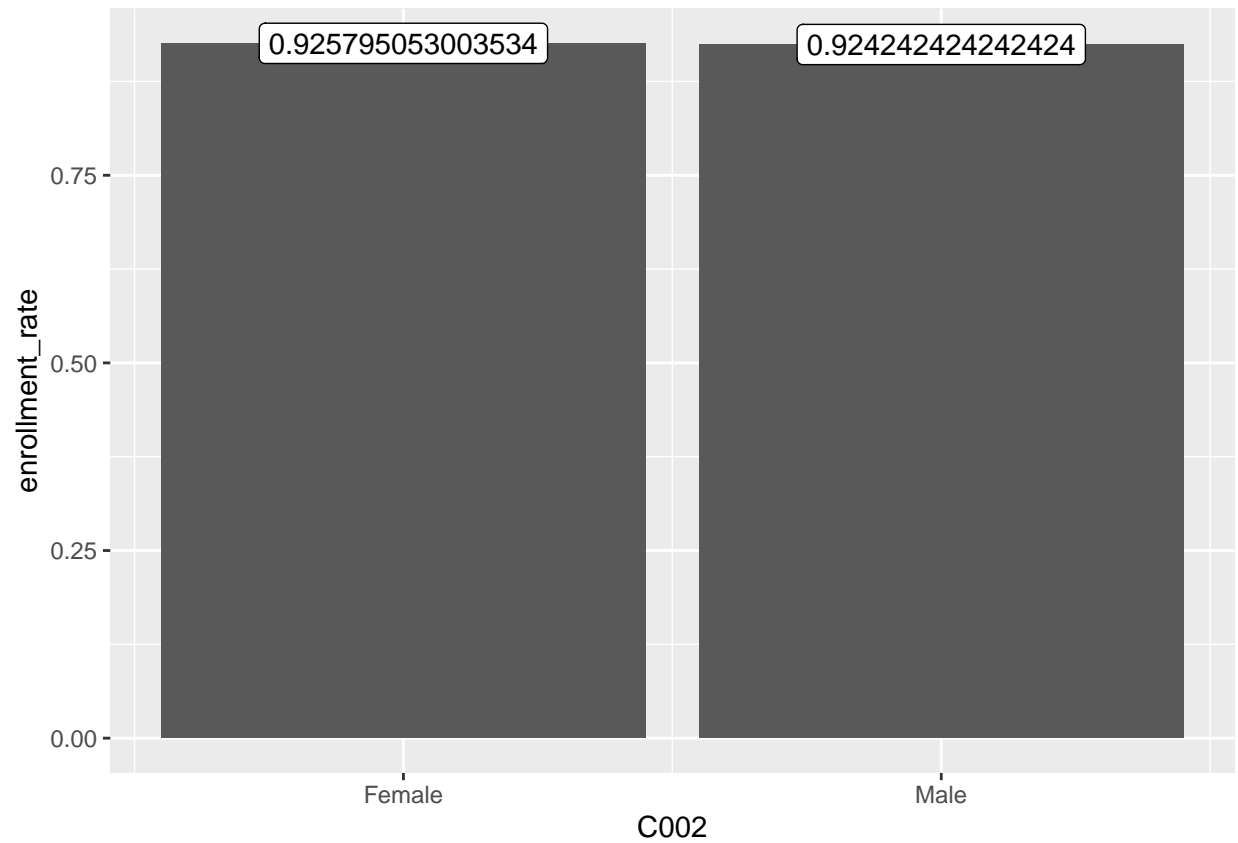
```r
child %>%
  group_by(DID) %>%
  mutate(avg = round(mean(C003 == 3), digits = 2)) %>%
  ungroup() %>%
  ggplot(aes(avg)) +
  geom_histogram() +
  facet_grid(~RID, labeller = labeller(RID = RegionName)) +
  labs(title = "Current Enrollment Rate by Region")
```

Comparison between Other Region

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Current Enrollment Rate by Region
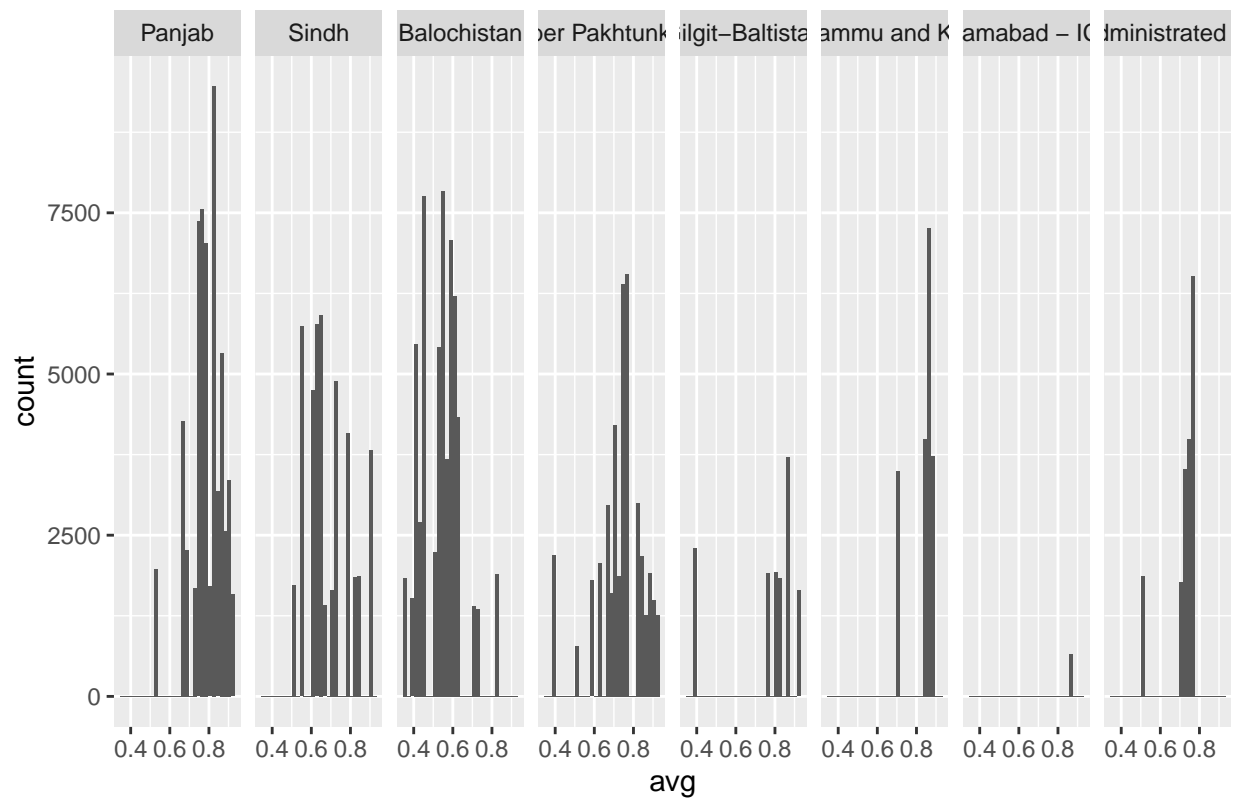
Within Gilgit-Baltistan Within Gilgit-Baltistan, Hunza is outperforming.

```
child %>%
  filter(RID == 6) %>%
  group_by(DID) %>%
  mutate(Current_Enrollment_Rate = mean(C003 == 3)) %>%
  ggplot(aes(DID, Current_Enrollment_Rate)) +
  geom_count() +
  scale_x_continuous(breaks = 260:266, labels = c("Gilgit", "Diamer", "Skardu", "Ghanshe", "Astore", "Gl
```