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

#### Packages Used

library(tidyverse)  
library(ggplot2)

#### Data Installation

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")

### Exploration

#### Checking Samplesizes

length(unique(child$CID))

## [1] 255196

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

child %>%   
 filter(DID == 266) %>%   
 summarize(N\_hunza = length(unique(CID)))

## N\_hunza  
## 1 1641

The samplesize of Hunza alone is 1641.

### Exploration in Hunza

#### Gender Proportion

child %>%   
 filter(DID == 266) %>%   
 summarize(gender\_proportion = mean(C002))

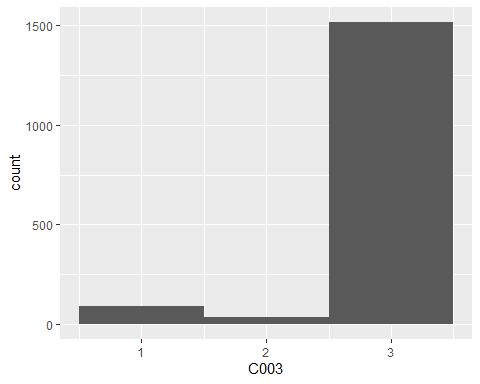
## gender\_proportion  
## 1 -0.5173675

-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

child %>%   
 filter(DID == 266) %>%   
 ggplot(aes(C003)) +  
 geom\_histogram(bins = 3)



#### Education Status by Gender

child %>%   
 filter(DID == 266) %>%   
 ggplot(aes(C003)) +  
 geom\_histogram(bins = 3, binwidth = 1) +  
 facet\_grid(~C002, labeller = labeller(C002 = Gender))

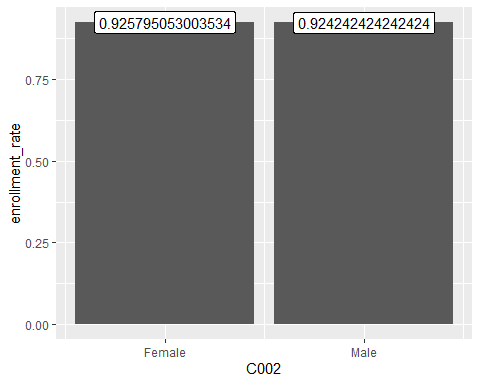


Both genders look pretty good interms of the absolute number of currently-enrolled-children

#### The Enrollment Rate by Gender

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)

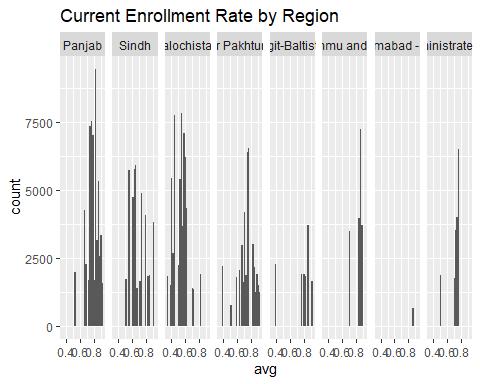


As a rate, both are doing pretty good

#### Comparison between Other Region

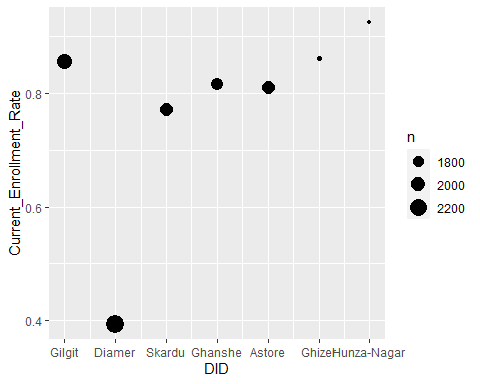
child %>%   
 group\_by(DID) %>%   
 mutate(avg = round(mean(C003 == 3), digits = 2)) %>%   
 ungroup() %>%   
 ggplot(aes(avg)) +  
 geom\_histogram() +  
 facet\_grid(~RID, labeller = labeller(RID = RegionName)) +  
 labs(title = "Current Enrollment Rate by Region")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



#### Within Gilgit-Baltistan

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", "Ghizer", "Hunza-Nagar"))



Within Gilgit-Baltistan, Hunza is outperforming.