CGDV Data Visualization Challenge

Team AU

April 12, 2019

Introduction

We are looking at the Availability of Water in Ethiopia. We support the key message:

"The time burden of collection is greater for those using unimproved sources, female and younger members of the household and residents of rural areas."

We extracted necessary data from both table 7 and 8.

Load Dataset

```
library(tidyverse)

Gender_of_Collector <- read_csv("Gender of Collector.csv")
names(Gender_of_Collector) <- c("Gender", "Country", "Rural", "Urban")

Type_of_Water_Source <- read_csv("Type of Water Source.csv")
names(Type_of_Water_Source) <- c("Quality", "OnPremises",   "1-30minutes",
"31-60minutes", "Over60minutes",    "Total")
```

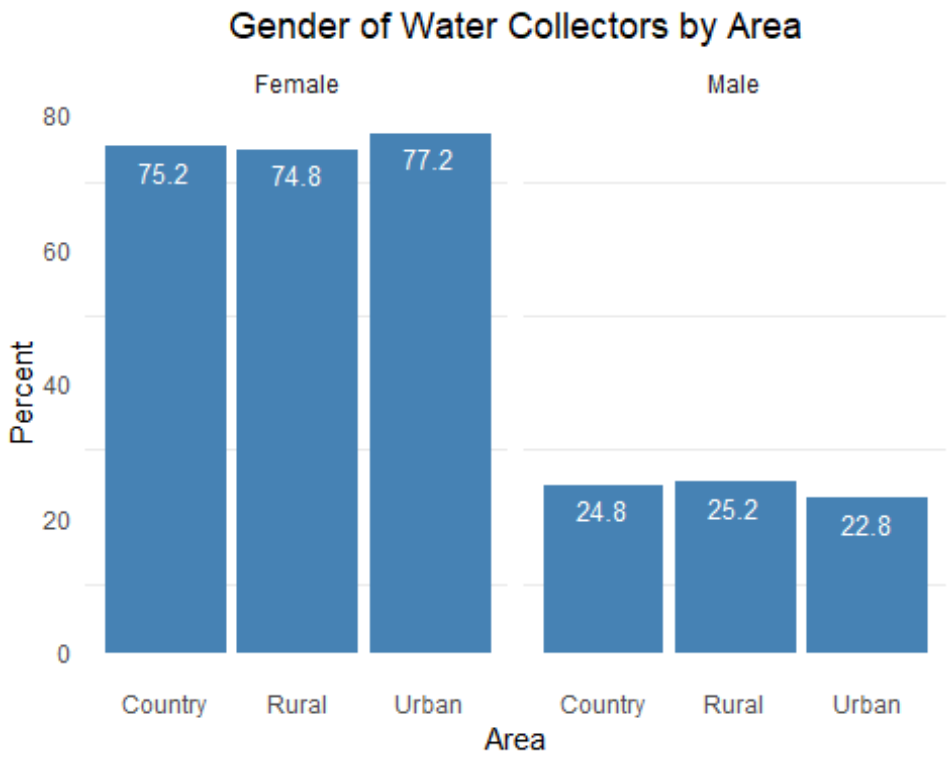
Gender of Water Collectors by Area

The time burden of collection is greater female in all areas of the country, with Urban area actually showing a slightly larger percentage of Female collecting water for the family.

```
knitr::kable(Gender_of_Collector)
```

Gender	Country	Rural	Urban
Male	24.8	25.2	22.8
Female	75.2	74.8	77.2
Total	100.0	100.0	100.0

```
Gender_of_Collector %>% gather(Area, Percent, 2:4) %>% #We transposed the
data to make visualiation easier.
  filter(Gender != "Total") %>%
  ggplot(aes(x = Area, y = Percent)) +
  geom_bar(stat = "identity",  fill = "steelblue", color = "steelblue") +
  facet_wrap(~Gender) +
  ggtitle("Gender of Water Collectors by Area")+
  geom_text(aes(label=Percent), vjust=1.6, color="white", size=3.5)+
  theme_minimal() +
  theme(panel.grid.major = element_blank(),
        plot.title = element_text(hjust = 0.5))
```



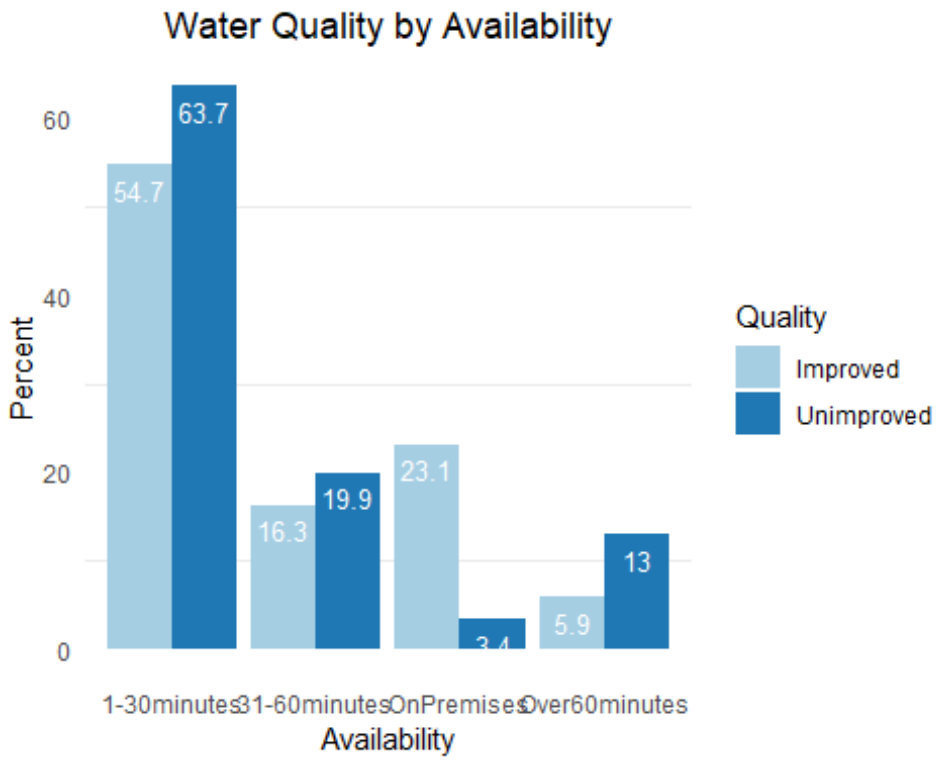
Water Quality by Availability

From the visualization below, it is clear that improved water sources are more likely to be available on premise than unimproved water (23.1% versus 3.4%), suggesting that the time burden of collection is greater for those using unimproved sources.

```
knitr::kable(Type_of_Water_Source)
```

Quality	OnPremises	1-30minutes	31-60minutes	Over60minutes	Total
Improved	23.1	54.7	16.3	5.9	100
Unimproved	3.4	63.7	19.9	13.0	100

```
Type_of_Water_Source %>% gather(Availability, Percent, 2:5) %>%
  ggplot(aes(x = Availability, y = Percent, fill = Quality)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_fill_brewer(palette = "Paired") +
  scale_color_brewer(palette = "Paired") +
  ggtitle("Water Quality by Availability") +
  geom_text(aes(label=Percent), vjust=1.6, color="white", size=3.5, position
= position_dodge(0.9))+
  theme_minimal() +
  theme(panel.grid.major = element_blank(),
        plot.title = element_text(hjust = 0.5))
```



Conclusion

Our barcharts support the key message that the time burden of collection is greater for those using unimproved sources and females.