

# Rural Urban dynamics (DTE Datathon)

Group 7

2023-09-26

## Introduction

This analysis is intended for the DTE datathon, which aims to cultivate a data-driven, innovative, and collaborative environment for creatively addressing societal issues and producing actionable solutions.

```
library(tidyverse)
library(gridExtra)
library(readxl)
library(reshape2)
library(janitor)
library(ggthemes)

library(knitr)
# library(kableExtra)

df <- data.frame(
  Details = c("Time Frame", "Deadline", "Software/Tools"),
  values = c("1 Week", "28th September", "R - studio")
)
df %>% kable(caption = "The rules of the datathon")
```

Table 1: The rules of the datathon

Details	values
Time Frame	1 Week
Deadline	28th September
Software/Tools	R - studio

**The objective:** The objective of this challenge is to analyze open data sets in Kenya using real-world data. You are therefore to explore, visualize, and draw insights from the provided dataset to provide better insights to different stakeholders on how urbanization is impacting different aspects of Kenyan society. Participants should work in teams and are encouraged to utilize technologies such as big data, machine learning, and artificial Intelligence that train test, and evaluate multiple data sets to uncover innovative solutions.

```
colors <- c('#BC5308', '#FFECD1', '#C5CAB8', '#FF7D00', '#8AA79F', '#FFB569', '#15616D', '#001524')

df <- read_excel("Dataset/Dataset.xlsx")
```

```
kenya <- df
kenya1 <- df
df %>% select(1:7) %>% head() %>% knitr::kable(caption = "A sample of the data")
```

Table 2: A sample of the data

COUNTY	County	Rural_ppn	Urban_ppn	Total Population	Urban/Total	Status
Baringo	Baringo	591474	75289	666763	0.1129172	Rural
Bomet	Bomet	847718	27971	875689	0.0319417	Rural
Bungoma	Bungoma	1480458	190112	1670570	0.1138007	Rural
Busia	Busia	779928	113753	893681	0.1272859	Rural
Elgeiyo-	Elgeiyo-	433901	20579	454480	0.0452803	Rural
Marakwet	Marakwet					
Embu	Embu	532675	75924	608599	0.1247521	Rural

## data

As per the rules, we sourced the data from KNBS on the following website: KNBS Data.

Additionally, we obtained supplementary data from the following website: Kenya County Fact Sheets Report (PDF).

The collected data encompasses a wide range of aspects of Kenyan life, including population, GDP, infrastructure, education, healthcare, and employment.

## data sourcing and cleaning

The data collection process was primarily manual, and it followed the following general steps:

1. **Data Compilation:** We collaborated in Google Sheets to collect various data points simultaneously, gradually building up the dataset we needed.
2. **Data Cleaning:** We employed a manual query system to identify and address any anomalies or outliers in the dataset. This process involved thorough cross-checking to ensure data accuracy.
3. **Feature Engineering:** We created additional columns in the dataset, such as:
  - Total Population:  $totalPopulation = femalePopulation + malePopulation$
  - Urban-Rural Classification:  $urbanRural = \frac{urbanPopulation}{totalPopulation}$

Based on the urban-rural classification, we categorized certain counties as either rural or urban. Counties were classified as fully urban if the urban population exceeded 40% of the total population.

## Rural-Urban population dynamics

## Rural-Urban health dynamics

## Population distribution

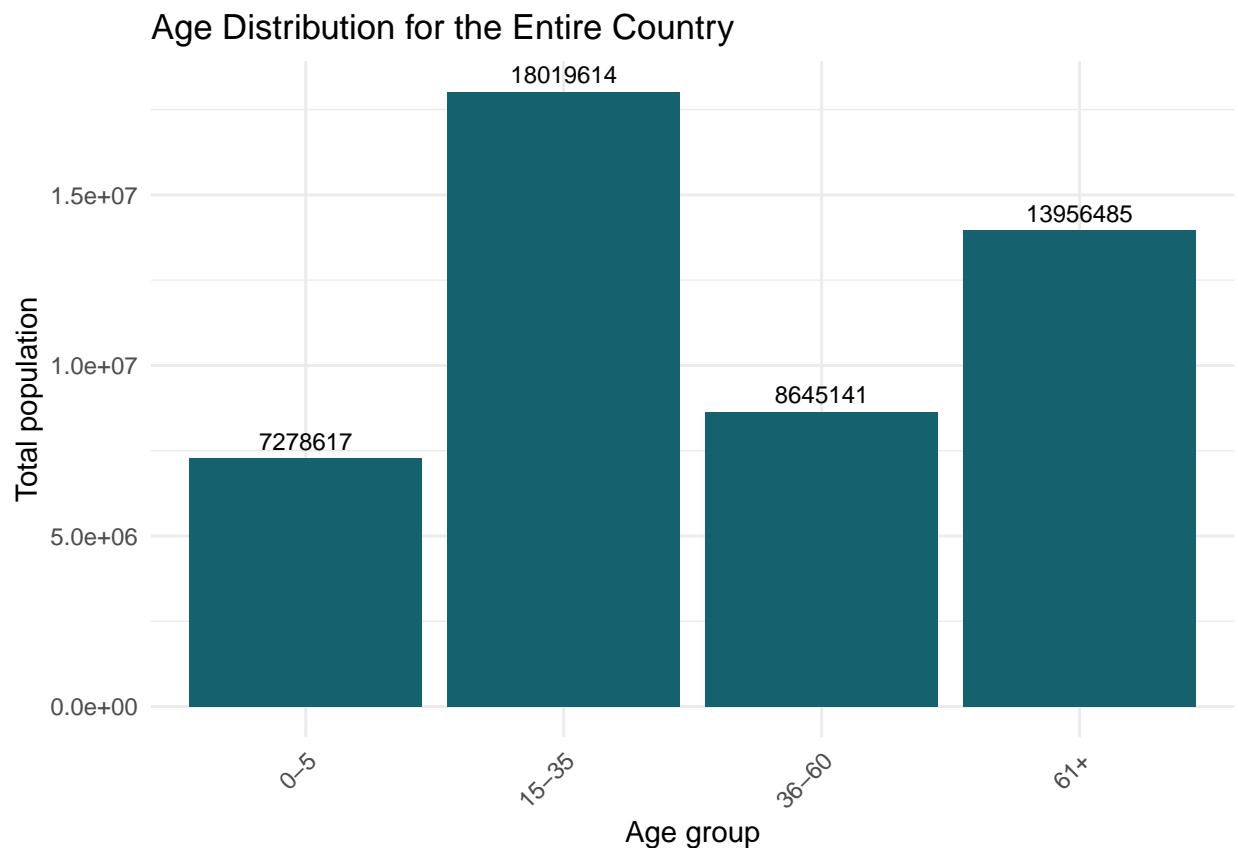
```

data<-read_excel("Dataset.xlsx",sheet="population distribution")

data$Cohorts<-as.factor(data$Cohorts)
data$Cohorts<-ifelse(data$Cohorts=="0-5", "0-5",
                     ifelse(data$Cohorts=="45091", "6-14",
                             ifelse(data$Cohorts=="15-35", "15-35",
                                     ifelse(data$Cohorts=="36-60", "36-60", "61+"))))
data$Cohorts <- factor(data$Cohorts, levels = c("0-5", "6-14", "15-35", "36-60", "61+"))
library(ggplot2)

# Create a histogram of age cohort distribution
data %>%
  group_by(Cohorts) %>%
  summarise(Totals=sum(Total)) %>%
ggplot( aes(x = Cohorts, y = Totals)) +
  geom_bar(stat = "identity", fill = '#15616D') +
  geom_text(aes(label = Totals), vjust = -0.5, size = 3) + # Add text labels above the bars
  labs(x = "Age group", y = "Total population", title = "Age Distribution for the Entire Country") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```

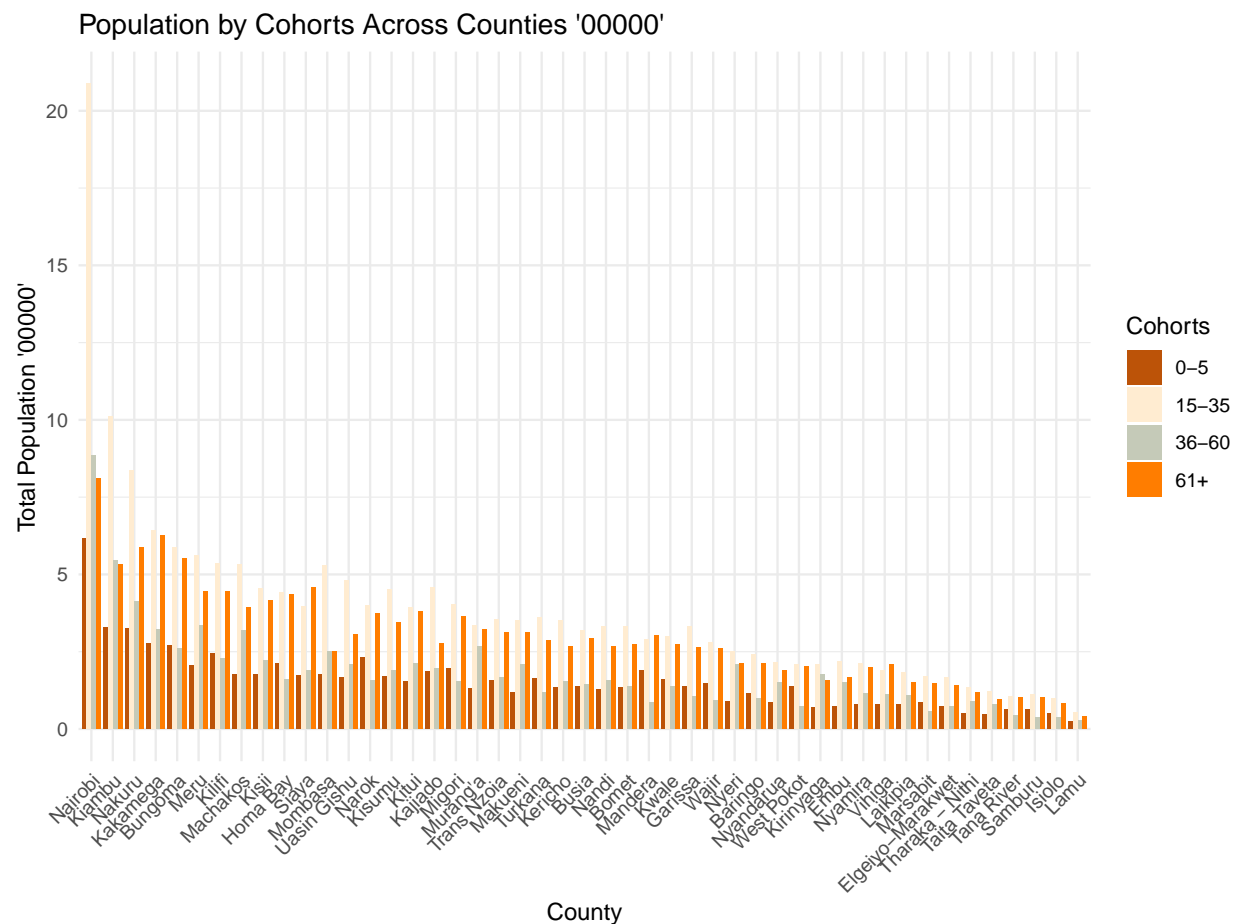


Persons aged 15 to 35 years are majority indicating the high labour force and the least population in the country are those above 60 years.

## The distribution of age groups in each county

```
cohort_summary <- data %>%
  group_by(County, Cohorts) %>%
  summarise(Total_Population = round((sum(Total)/100000),3))

# Create the bar graph
ggplot(cohort_summary, aes(x = reorder(County, -Total_Population), y = Total_Population, fill = Cohorts)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(x = "County", y = "Total Population '00000'", title = "Population by Cohorts Across Counties '00000'") +
  scale_fill_discrete() +
  theme_minimal() +
  scale_fill_manual(values = colors) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Majority of the counties had a population less than one million across all the age groups.

## Population distribution for each county by each gender

```
data_long <- data %>%
  pivot_longer(cols = -c("County", "Cohorts", "Total"), names_to = "gender", values_to = "population")
```

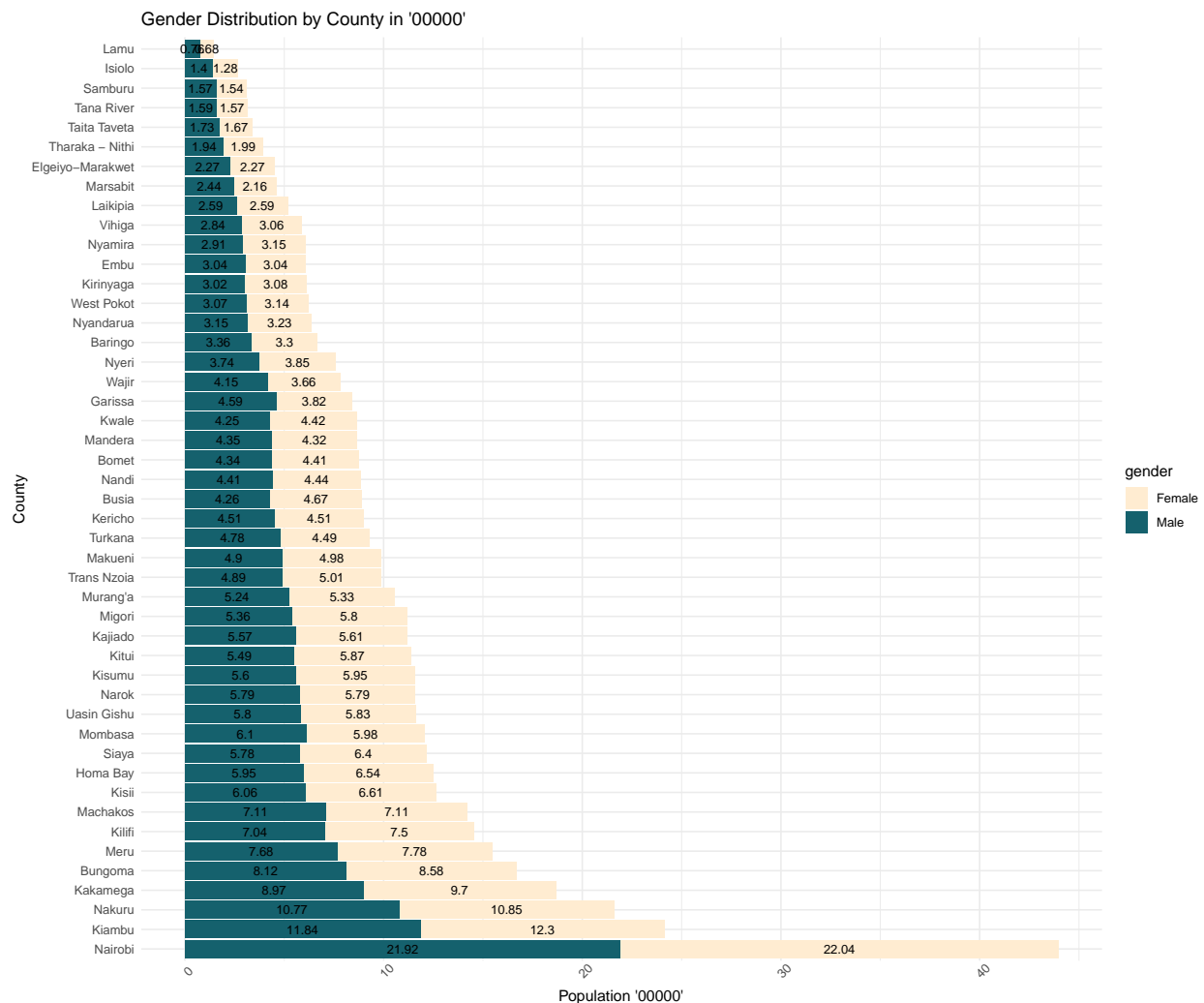
```

# Calculate Total population per County
total_population <- data_long %>%
  group_by(County) %>%
  summarise(Total_Population = sum(population))

# Reorder the levels of County based on Total population (descending order)
data_long$County <- factor(data_long$County, levels = total_population$County[order(-total_population$Total_Population)])

data_long %>%
  group_by(County, gender) %>%
  summarise(Totals = round((sum(population) / 100000), 2)) %>%
  ggplot(aes(x = County, y = Totals, fill = gender)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = Totals), position = position_stack(vjust = 0.5), size = 3) +
  labs(y = "Population '00000'", x = "County", title = "Gender Distribution by County in '00000'") +
  scale_fill_manual(values = c("Female" = "#FFECD1", "Male" = "#15616D")) +
  theme_minimal() +
  coord_flip() +
  theme(axis.text.x = element_text(angle = 45, hjust = 0.5))

```



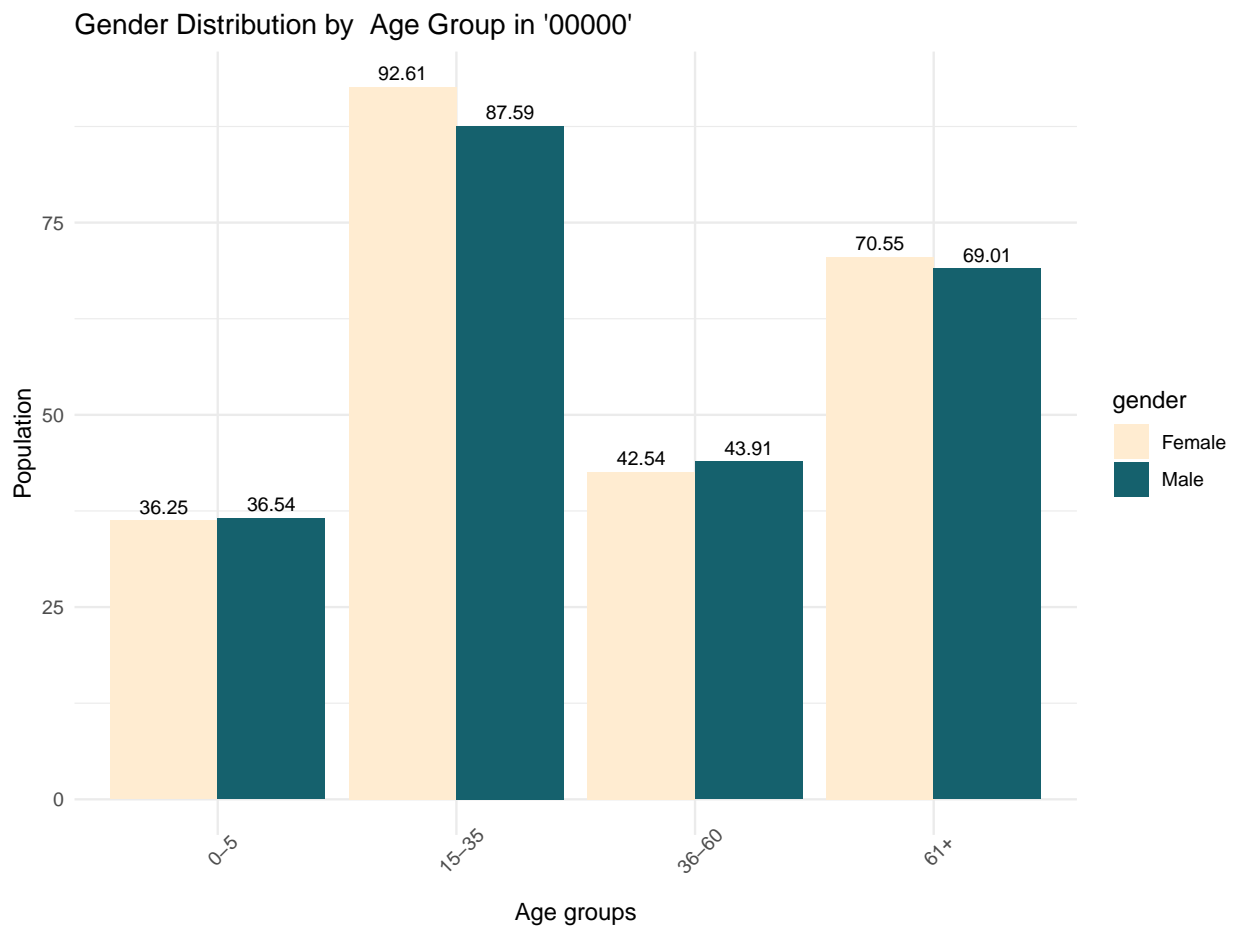
Nairobi county has the highest population by both gender levels and Lamu county has the lowest population with more males than females.

### Gender distribution across the age groups in the population.

*# Assuming you have already loaded and transformed your data into 'data\_long' as described in your code*

```
data_long %>%
  group_by( Cohorts, gender) %>%
  summarise(Totals = round((sum(population) / 100000), 2)) %>%
  ggplot(aes(x = Cohorts, y = Totals, fill = gender)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_text(aes(label = Totals), position = position_dodge(width = 0.9), size = 3, vjust = -0.5) +
  labs(y = "Population", x = "Age groups", title = "Gender Distribution by Age Group in '00000'") +
  scale_fill_manual(values = c("Female" = "#FFECD1", "Male" = "#15616D")) +
  theme_minimal() +

  theme(axis.text.x = element_text(angle = 45, hjust = 0.5))
```



Female gender in the population were more in age groups 15to 35 years and above 60 years while contrast to male gender in the other age groups.

## Population mean distribution in Rural and Urban in Kenya

```
stats=df %>% summarise(`Rural mean population`=mean(Rural_ppn),
                        `Rural SD`=sd(Rural_ppn),
                        `Rural IQR`=IQR(Rural_ppn),
                        `Urban mean population`=mean(Urban_ppn),
                        `Urban SD`=sd(Urban_ppn),
                        `Urban IQR`=IQR(Urban_ppn))
stats %>% kable()
```

Rural mean population	Rural SD	Rural IQR	Urban mean population	Urban SD	Urban IQR
686186.6	376807.1	470351	315073.8	689484.4	112220.5

On average the rural distribution is (Mean=696932.3 persons ,SD=367022) and urban population is (Mean=315073.8,SD= 689484.4)

## Rural and urban population comparison for each county

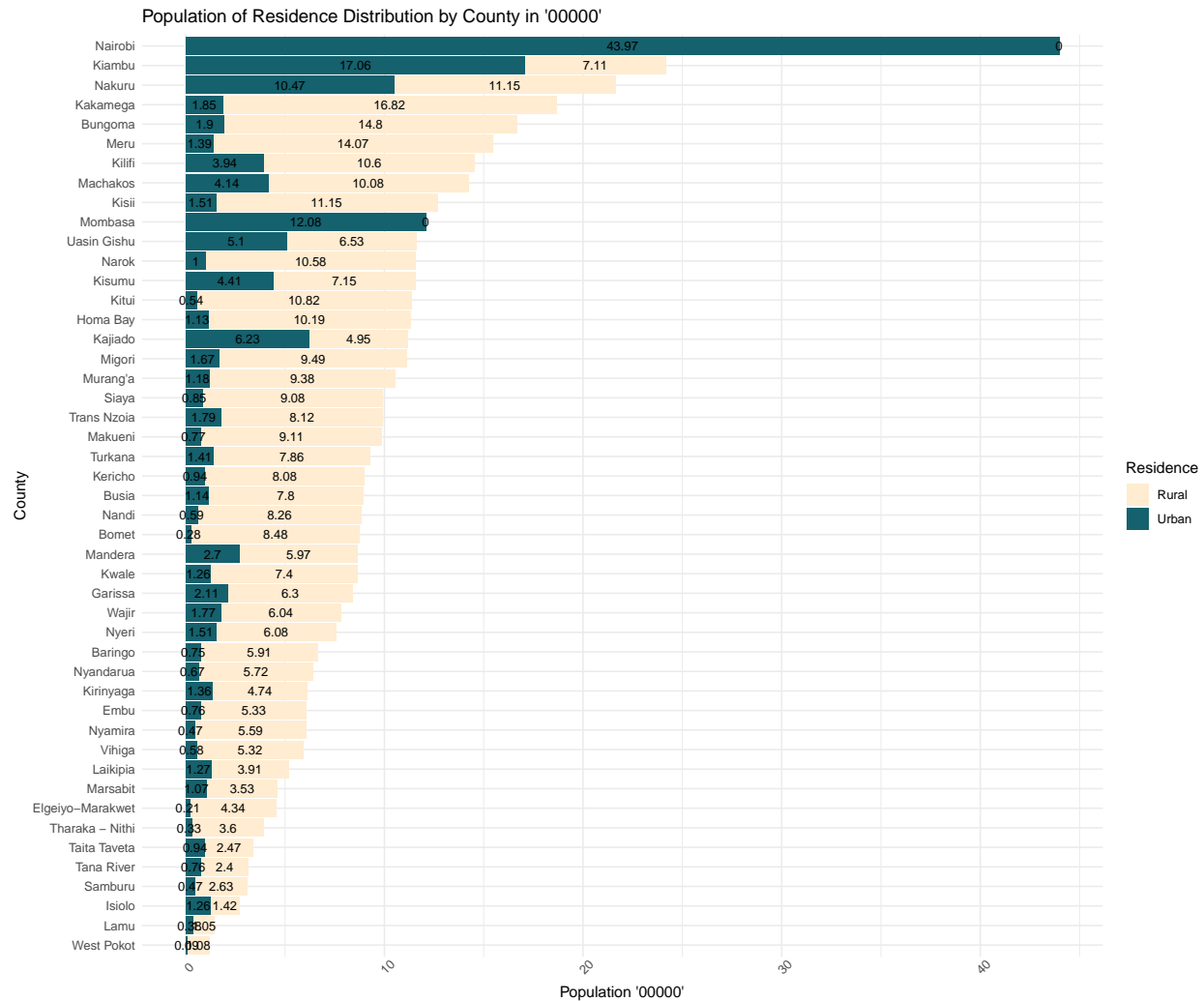
```
DataRural <- df %>%
  select(County, Rural_ppn, Urban_ppn) %>%
  pivot_longer(cols = c(Rural_ppn, Urban_ppn), names_to = "Residence", values_to = "Populations")

DataRural$Residence <- ifelse(DataRural$Residence == "Rural_ppn", "Rural", "Urban")
DataRural$Populations <- as.numeric(DataRural$Populations)

# Calculate Total population per County
total_population <- DataRural %>%
  group_by(County) %>%
  summarise(Total_Population = sum(Populations))

# Reorder the levels of County based on Total population (descending order)
DataRural$County <- factor(DataRural$County, levels = total_population$County[order(total_population$Total_Population)])

DataRural %>%
  group_by(County, Residence) %>%
  summarise(Totals = round((sum(Populations) / 100000), 2)) %>%
  ggplot(aes(x = County, y = Totals, fill = Residence)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = Totals), position = position_stack(vjust = 0.5), size = 3) +
  labs(y = "Population '00000'", x = "County", title = "Population of Residence Distribution by County") +
  scale_fill_manual(values = c("Rural" = "#FFECD1", "Urban" = "#15616D")) +
  theme_minimal() +
  coord_flip() +
  theme(axis.text.x = element_text(angle = 45, hjust = 0.5))
```



Mombasa and Nairobi had the residence being urban and Nakuru county accounted half of the population as urban and rural residence. While West pokot county accounted majority of its residence as rural.

```
library(leaflet)
library(sf)

# Load the shapefile
kenya_counties <- st_read("LYNN ANALYSIS/County.shp")
```

Reading layer 'County' from data source

```
'D:\Willy\DTE-Datathon\LYNN ANALYSIS\County.shp' using driver 'ESRI Shapefile'
Simple feature collection with 47 features and 8 fields
Geometry type: MULTIPOLYGON
Dimension: XY
Bounding box: xmin: 33.91182 ymin: -4.702271 xmax: 41.90626 ymax: 5.430648
Geodetic CRS: WGS 84
```

```
# Merge data by county
merged_data <- merge(kenya_counties, kenya, by.x = "COUNTY", by.y = "COUNTY")
```



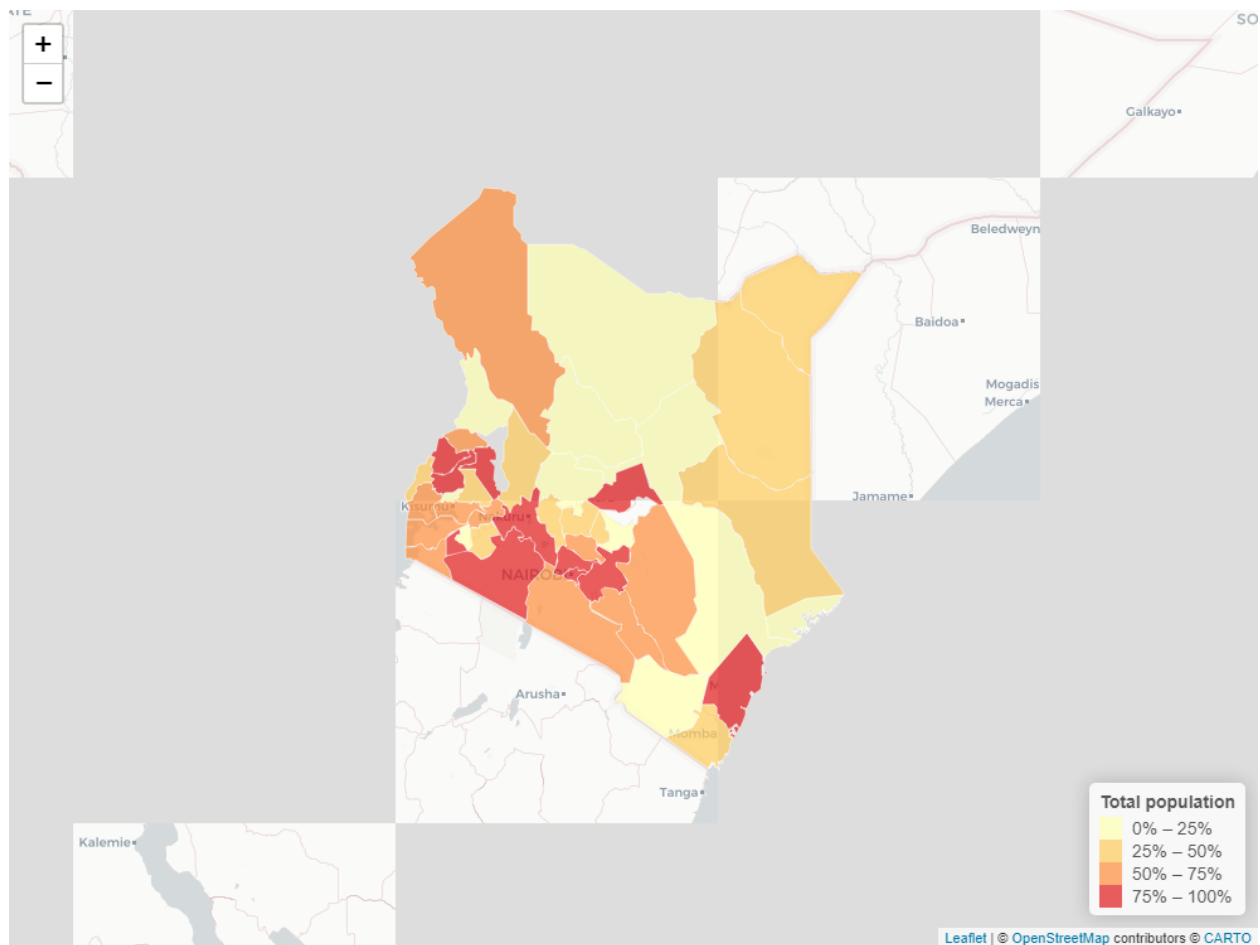
```

# Create a leaflet map
#install.packages("leaflet")
#remotes::install_github("rstudio/webshot2")
library(leaflet)
library(webshot2)
kenya_map <- leaflet(merged_data) %>%
  addProviderTiles("CartoDB.Positron") %>%
  addPolygons(
    fillColor = ~colorQuantile("YlOrRd", merged_data$`Total Population`)(merged_data$`Total Population`),
    fillOpacity = 0.7,
    color = "white",
    weight = 1,
    label = ~paste(merged_data$COUNTY, "<br>", "Total Population:", merged_data$`Total Population`, "people"),
    highlight = highlightOptions(
      weight = 3,
      color = "#666",
      fillOpacity = 0.7,
      bringToFront = TRUE
    ),
    labelOptions = labelOptions(
      style = list("font-weight" = "normal", padding = "3px 8px"),
      textsize = "15px"
    )
  ) %>%
  addLegend(
    "bottomright",
    pal = colorQuantile("YlOrRd", merged_data$`Total Population`),
    values = ~merged_data$`Total Population`,
    title = "Total population",
    opacity = 0.7
  )

# Display the map
htmlwidgets::saveWidget(kenya_map, "kenya_map.html")

# Take a screenshot and save it as an image
webshot2::webshot("kenya_map.html", "kenya_map.png")

```



#Health Analysis

```
# Filter and summarize the data to count level 5/6 hospitals by county
level_5_6_hospitals <- kenya %>%
  filter(`Health_LV5/6` > 0)%>%
  select(COUNTY, `Health_LV5/6`)
#group_by(COUNTY) %>%
#summarise(Count_Level_5_6_Hospitals = sum(Health_LV5/6))

# Create a table using kable
#kable(level_5_6_hospitals, caption = "Count of Level 5/6 Hospitals by County")
```

```
# Load the required libraries
library(ggplot2)

# Filter the data to include only rural counties
kenya_rural <- kenya %>%
  filter(Status == "Rural") %>%
  arrange(Health_LV4)

# Create the bar chart
fig1 <- ggplot(kenya_rural, aes(x = reorder(COUNTY, -Health_LV4), y = Health_LV4, fill = "#BC5308")) +
  geom_bar(stat = "identity", show.legend = FALSE) +
  theme_minimal() +
```

```

labs(x = "County", y = "Number of Level 4 Hospitals") +
scale_fill_discrete(name = "County") +
ggtitle("Distribution of Level 4 Hospitals by Rural Counties") +
coord_flip() +
geom_text(aes(label = Health_LV4), vjust = 0.5, hjust = 1, size = 4, color = colors[7])+
theme_hc()

```

*# Filter the data to include only rural counties*

```

kenya_rural <- kenya %>%
  filter(Status == "Urban") %>%
  arrange(Health_LV4)

```

*# Create the bar chart*

```

fig2<-ggplot(kenya_rural, aes(x = reorder(COUNTY, -Health_LV4), y = Health_LV4)) +
  geom_bar(stat = "identity", show.legend = FALSE,fill=colors[1]) +
  theme_minimal() +
  labs(x = "County", y = "Number of Level 4 Hospitals") +
  scale_fill_discrete(name = "County") +
  ggtitle("Distribution of Level 4 Hospitals by Urban Counties") +
  coord_flip()+
  geom_text(aes(label = Health_LV4), vjust = 0.5, hjust = 1, size = 8, color = colors[7])+
  theme_hc()

```

*# Filter the data to include only rural counties*

```

kenya_rural <- kenya %>%
  filter(Status == "Rural") %>%
  arrange(Health_LV3)

```

*# Create the bar chart*

```

fig3<-ggplot(kenya_rural, aes(x = reorder(COUNTY, -Health_LV3), y = Health_LV3, fill = colors[1])) +
  geom_bar(stat = "identity", show.legend = FALSE) +
  theme_minimal() +
  labs(x = "County", y = "Number of Level 3 Hospitals") +
  scale_fill_discrete(name = "County") +
  ggtitle("Distribution of Level 3 Hospitals by Rural Counties") +
  coord_flip()+
  geom_text(aes(label = Health_LV3), vjust = 0.5, hjust = -0.1, size = 4, color = colors[7])+
  theme_hc()

```

*# Filter the data to include only rural counties*

```

kenya_rural <- kenya %>%
  filter(Status == "Urban") %>%
  arrange(Health_LV3)

```

*# Create the bar chart*

```

fig4<-ggplot(kenya_rural, aes(x = reorder(COUNTY, -Health_LV3), y = Health_LV3)) +
  geom_bar(stat = "identity", show.legend = FALSE,fill=colors[7]) +
  theme_minimal() +
  labs(x = "County", y = "Number of Level 3 Hospitals") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_fill_discrete(name = "County") +
  ggtitle("Distribution of Level 3 Hospitals by Urban Counties") +
  coord_flip()+

```

```
geom_text(aes(label = Health_LV3), vjust = 0.5, hjust = 1, size = 6, color = colors[1])+
theme_hc()
```

```
kenya_rural<-kenya%>%
  filter(Status=="Rural")%>%
  arrange(Health_LV2)

#create bar chart
fig5<-ggplot(kenya_rural, aes(x = reorder(COUNTY, -Health_LV2), y = Health_LV2, fill = colors[1])) +
  geom_bar(stat = "identity",show.legend = FALSE) +
  theme_minimal() +
  labs(x = "County", y = "Number of Level 2 Hospitals") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_fill_discrete(name = "County") +
  ggtitle("Distribution of Level 2 Hospitals by Rural Counties")+
  coord_flip()+
  geom_text(aes(label = Health_LV2), vjust = 0.5, hjust = 1, size = 3, color = colors[7])+
  theme_hc()
```

```
kenya_rural<-kenya%>%
  filter(Status=="Urban")%>%
  arrange(Health_LV2)

#create bar chart
fig6<-ggplot(kenya_rural, aes(x = reorder(COUNTY, -Health_LV2), y = Health_LV2)) +
  geom_bar(stat = "identity",show.legend = FALSE,fill=colors[1]) +
  theme_minimal() +
  labs(x = "County", y = "Number of Level 2 Hospitals") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_fill_discrete(name = "County") +
  ggtitle("Distribution of Level 2 Hospitals by Urban Counties")+
  coord_flip()+
  geom_text(aes(label = Health_LV2), vjust = 0.5, hjust = 1, size = 6, color = colors[7])+
  theme_hc()
```

```
kenya_rural<-kenya%>%
  filter(Status=="Rural")%>%
  arrange(`Bed density per 10000 ppn`)

#create bar chart
fig7<-ggplot(kenya_rural, aes(x = reorder(COUNTY, -`Bed density per 10000 ppn`), y = `Bed density per 10000 ppn`, fill="#15616D")) +
  geom_bar(stat = "identity",show.legend = FALSE) +
  theme_minimal() +
  labs(x = "County", y = "Bed density per 10000 ppn") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_fill_discrete(name = "County") +
  ggtitle("Distribution of Bed density per 10000 ppn by Rural Counties")+
  coord_flip()+
  geom_text(aes(label = `Bed density per 10000 ppn`), vjust = 0.5, hjust = 1, size = 4, color = colors[7])+
  theme_hc()
```

```
kenya_rural<-kenya%>%
  filter(Status=="Urban")%>%
  arrange(`Bed density per 10000 ppn`)

#create bar chart
fig8<-ggplot(kenya_rural, aes(x = reorder(COUNTY, -`Bed density per 10000 ppn`), y = `Bed density per 10000 ppn`)) +
  geom_bar(stat = "identity",show.legend = FALSE,fill=colors[7]) +
  theme_minimal() +
  labs(x = "County", y = "Bed density per 10000 ppn") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_fill_discrete(name = "County") +
  ggtitle("Distribution of Bed density per 10000 ppn by Urban Counties")+
  coord_flip()+
  scale_fill_manual(values = colors)+
  geom_text(aes(label = `Bed density per 10000 ppn`), vjust = 0.5, hjust = 1, size = 6, color = colors[7]) +
  theme_hc()
```

```
kenya_rural<-kenya%>%
  filter(Status=="Rural")%>%
  arrange(`Health worker per 10000 ppn`)

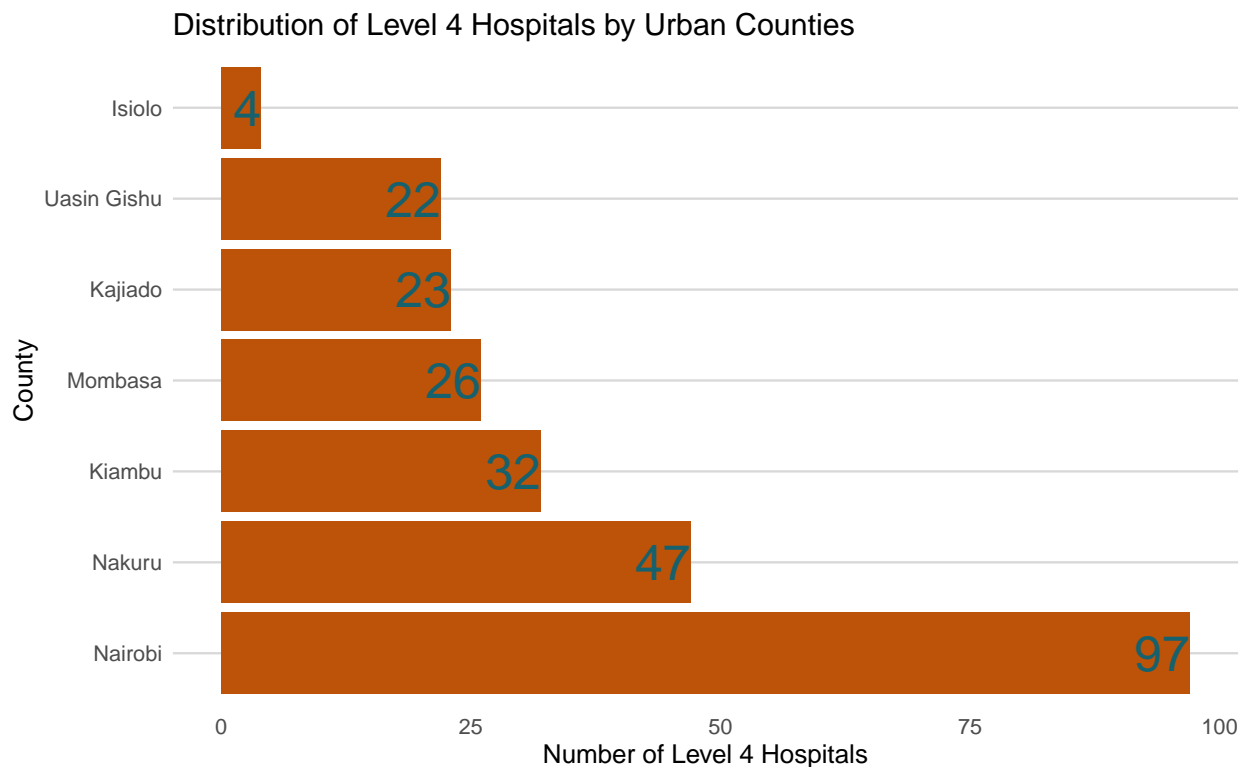
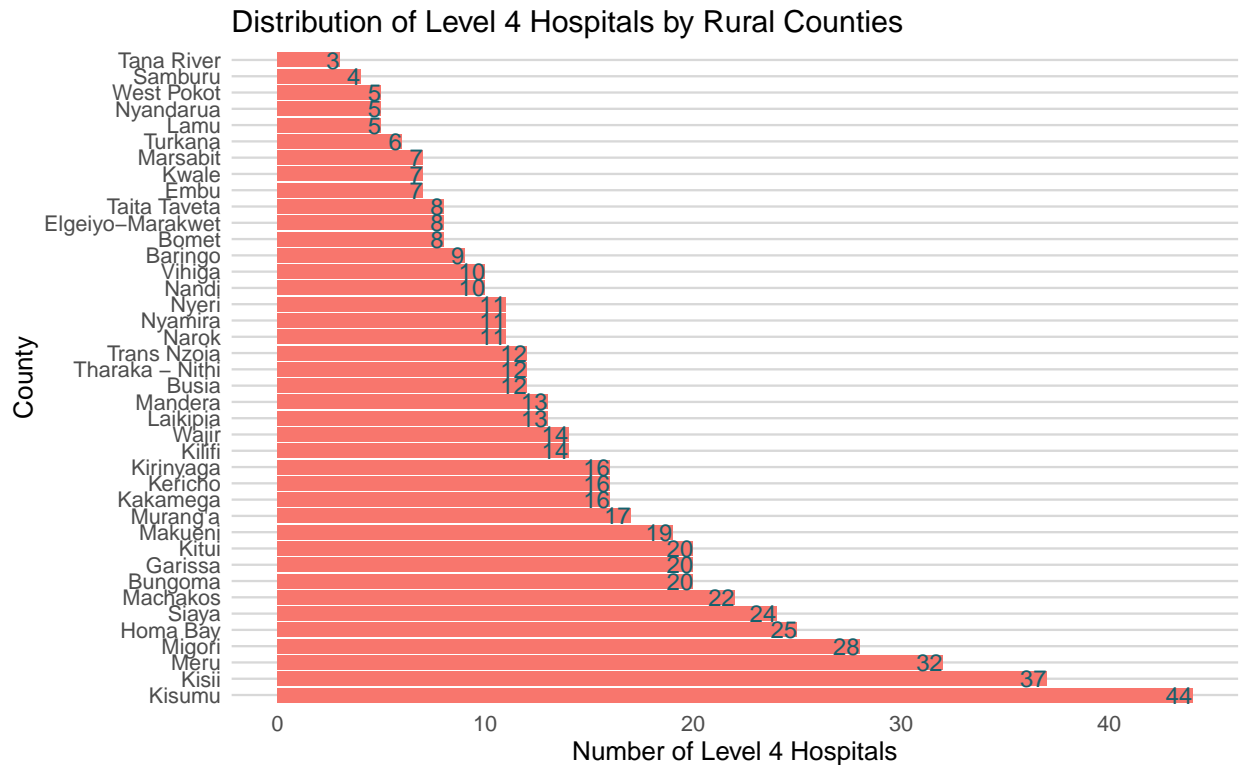
#create bar chart
fig9<-ggplot(kenya_rural, aes(x = reorder(COUNTY, -`Health worker per 10000 ppn`), y = `Health worker per 10000 ppn`)) +
  geom_bar(stat = "identity",show.legend = FALSE,fill="#15616D") +
  theme_minimal() +
  labs(x = "County", y = "Health worker per 10000 ppn") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_fill_discrete(name = "County") +
  ggtitle("Distribution of Health worker per 10000 ppn by Urban Counties")+
  coord_flip()+
  scale_fill_manual(values = colors)+
  geom_text(aes(label = `Health worker per 10000 ppn`), vjust = 0.5, hjust = 1, size = 4, color = colors[7]) +
  theme_hc()
```

```
kenya_rural<-kenya%>%
  filter(Status=="Urban")%>%
  arrange(`Health worker per 10000 ppn`)

#create bar chart
fig10<-ggplot(kenya_rural, aes(x = reorder(COUNTY, -`Health worker per 10000 ppn`), y = `Health worker per 10000 ppn`)) +
  geom_bar(stat = "identity",show.legend = FALSE,fill=colors[7]) +
  theme_minimal() +
  labs(x = "County", y = "Health worker per 10000 ppn") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_fill_discrete(name = "County") +
  ggtitle("Distribution of Health worker per 10000 ppn by Urban Counties")+
  coord_flip()+
  scale_fill_manual(values = colors)+
  geom_text(aes(label = `Health worker per 10000 ppn`), vjust = 0.5, hjust = 1, size = 6, color = colors[7]) +
  theme_hc()
```

## Bar charts showing Level 4 Hospital Distribution by counties

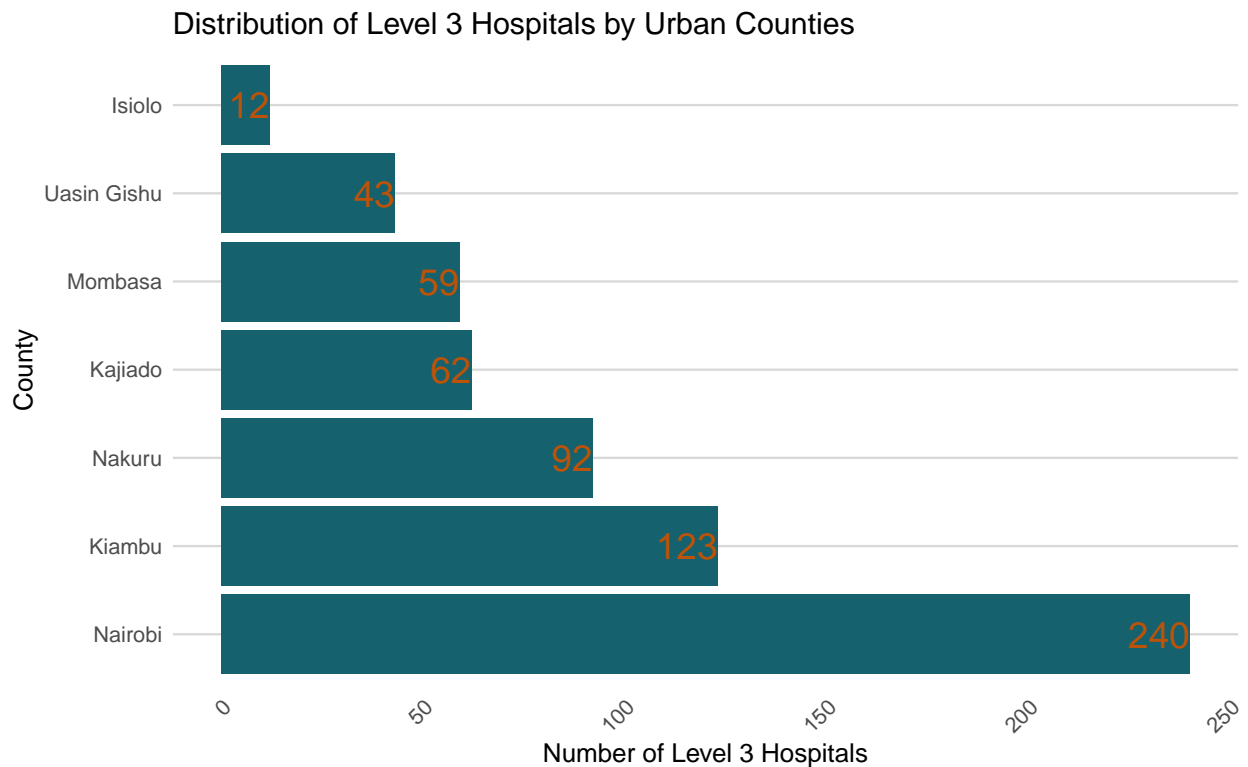
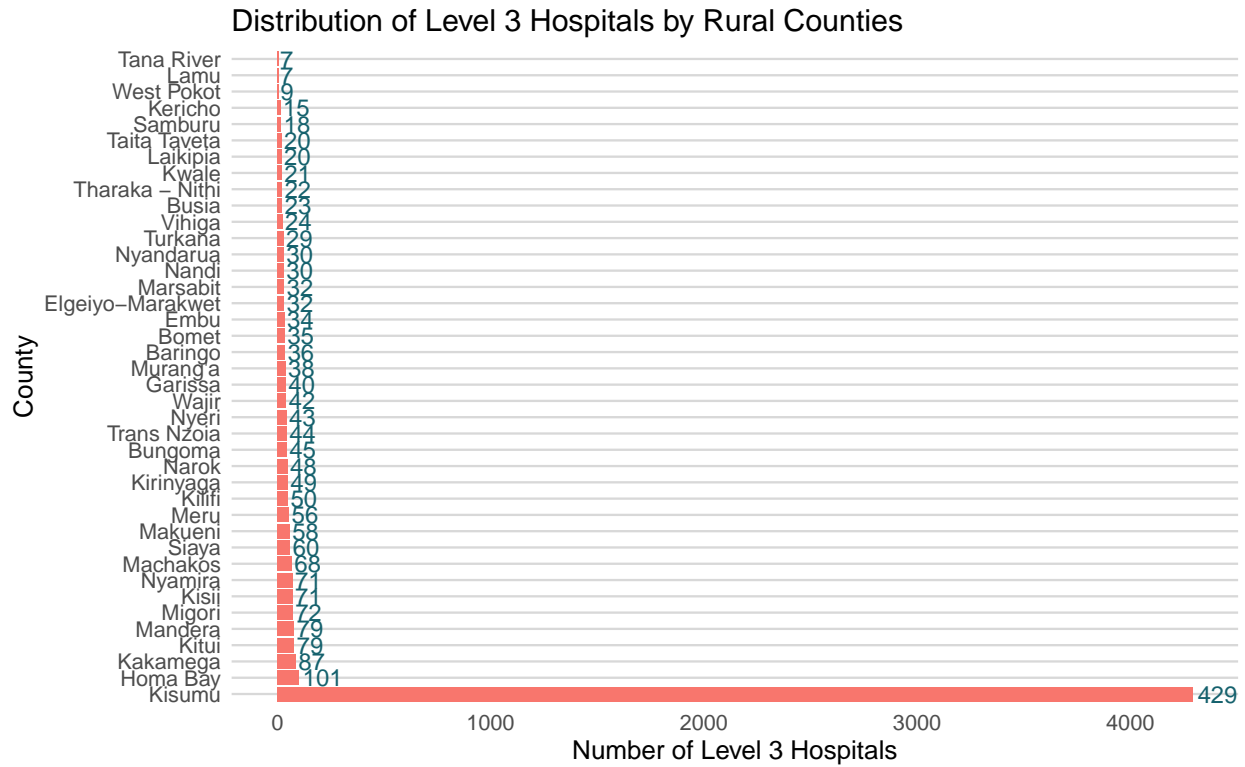
```
library(gridExtra)
grid.arrange(fig1,fig2,ncol=1)
```



Nairobi county accounted the most of level 4 hospitals among the urban centres and the least was Isiolo. In the rural category category Kisumu which is part of urban centre accounted most

## Bar charts showing Level 3 Hospital Distribution by counties

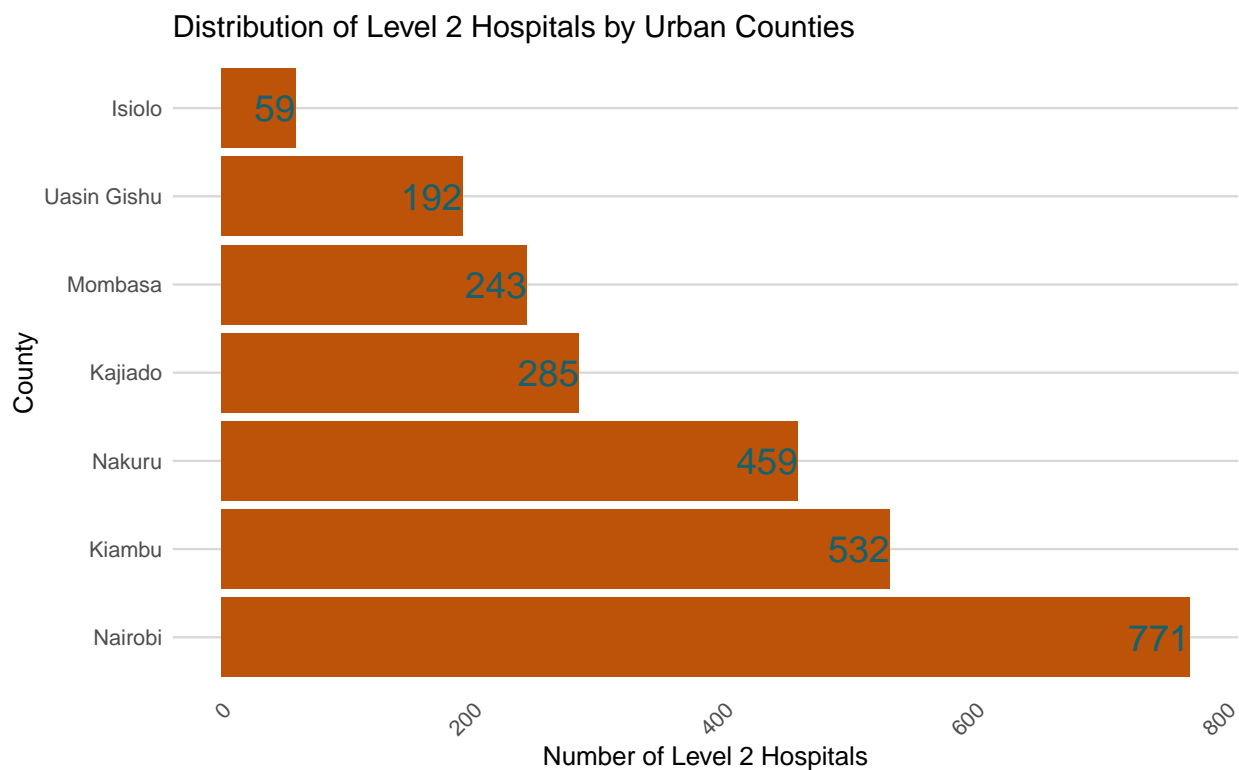
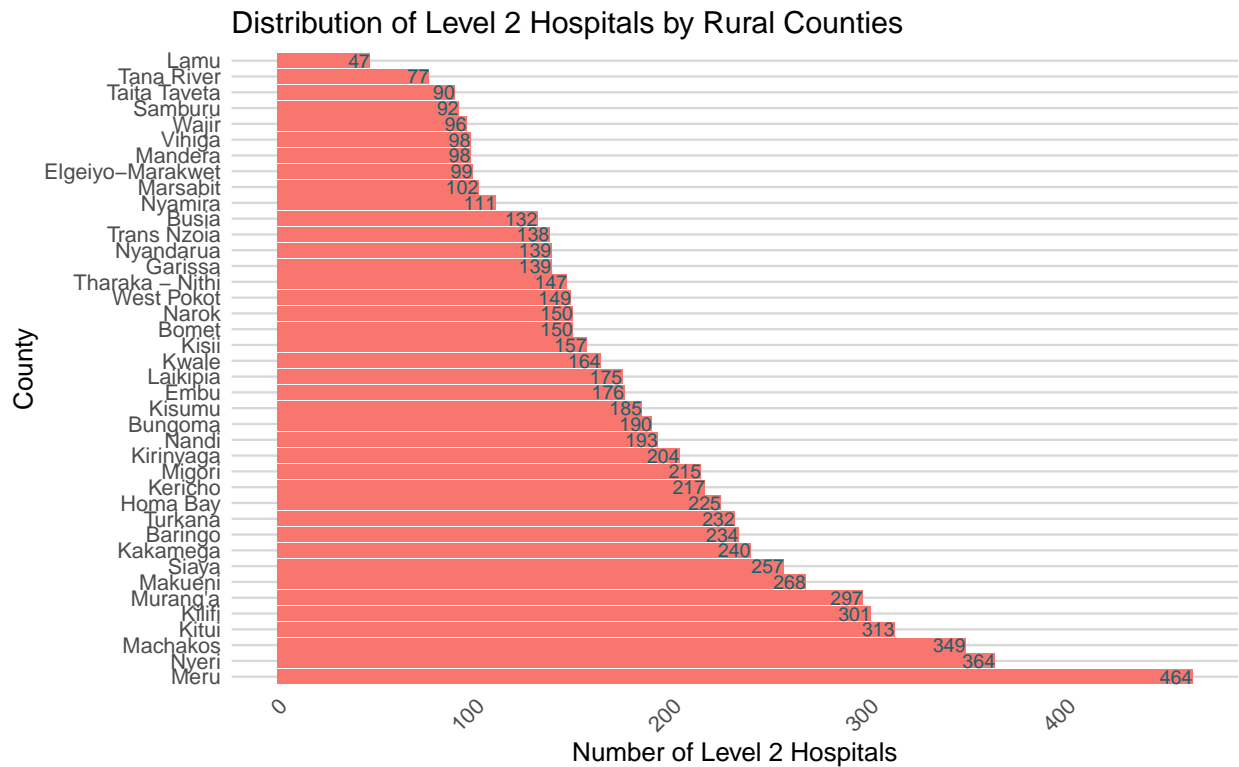
```
grid.arrange(fig3,fig4,ncol=1)
```





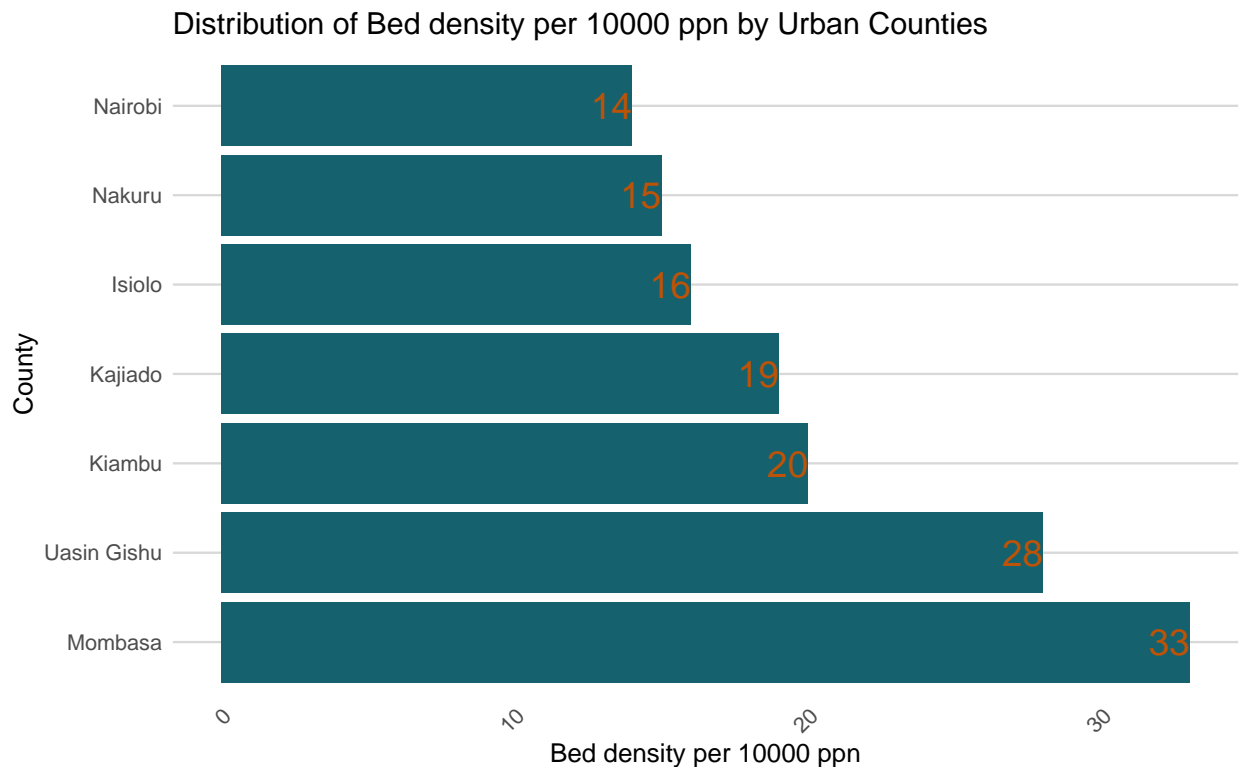
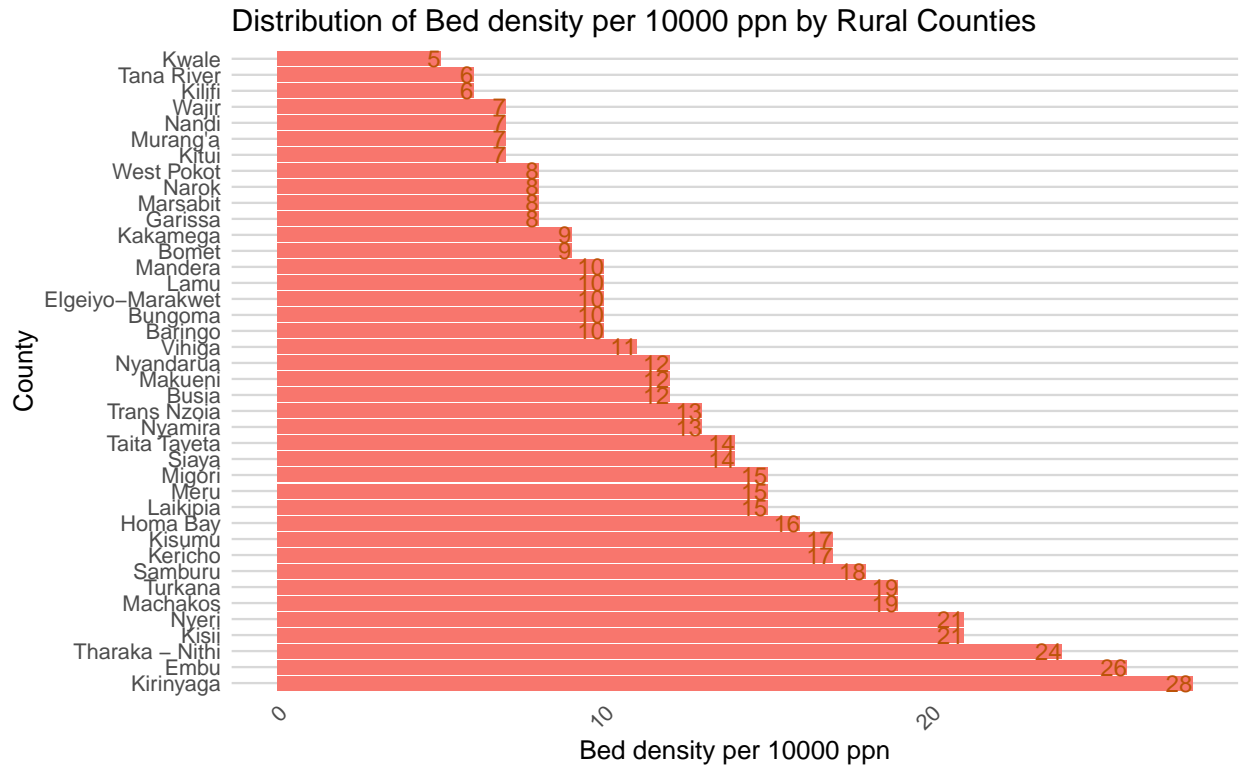
## Bar charts showing Level 2 Hospital Distribution by counties

```
grid.arrange(fig5,fig6,ncol=1)
```



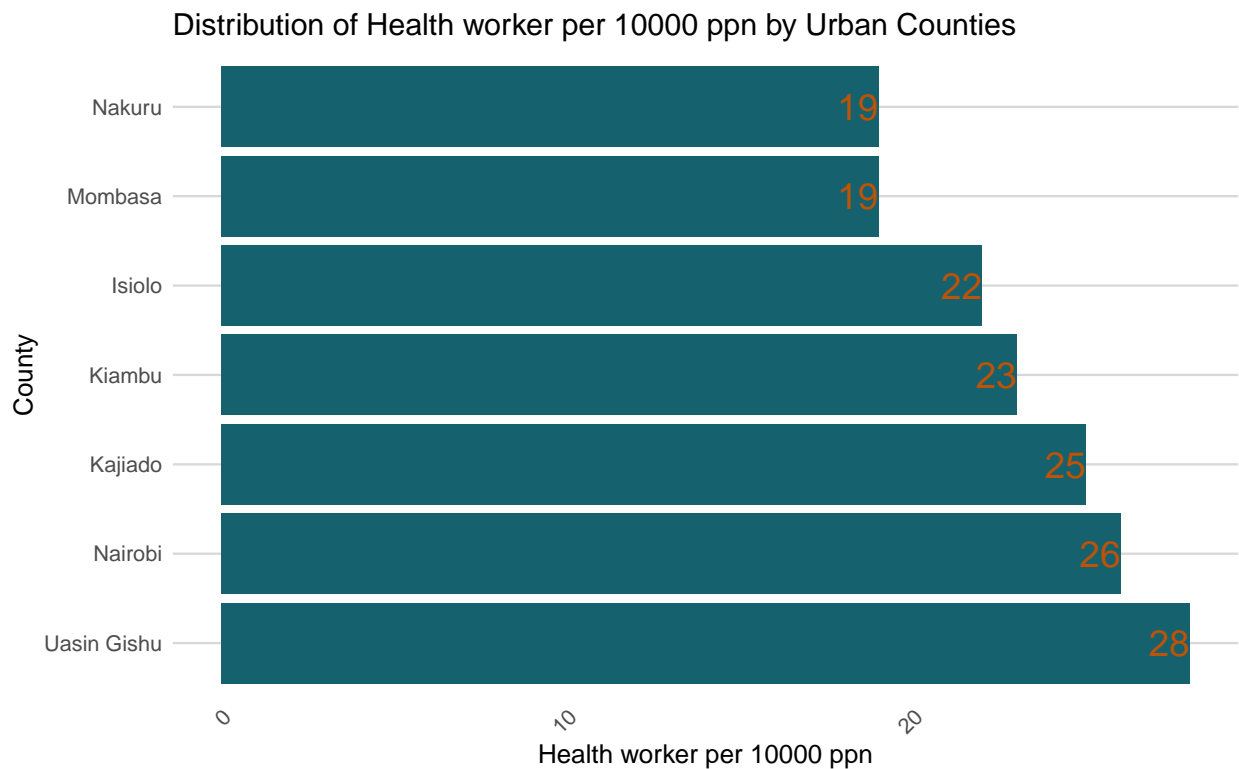
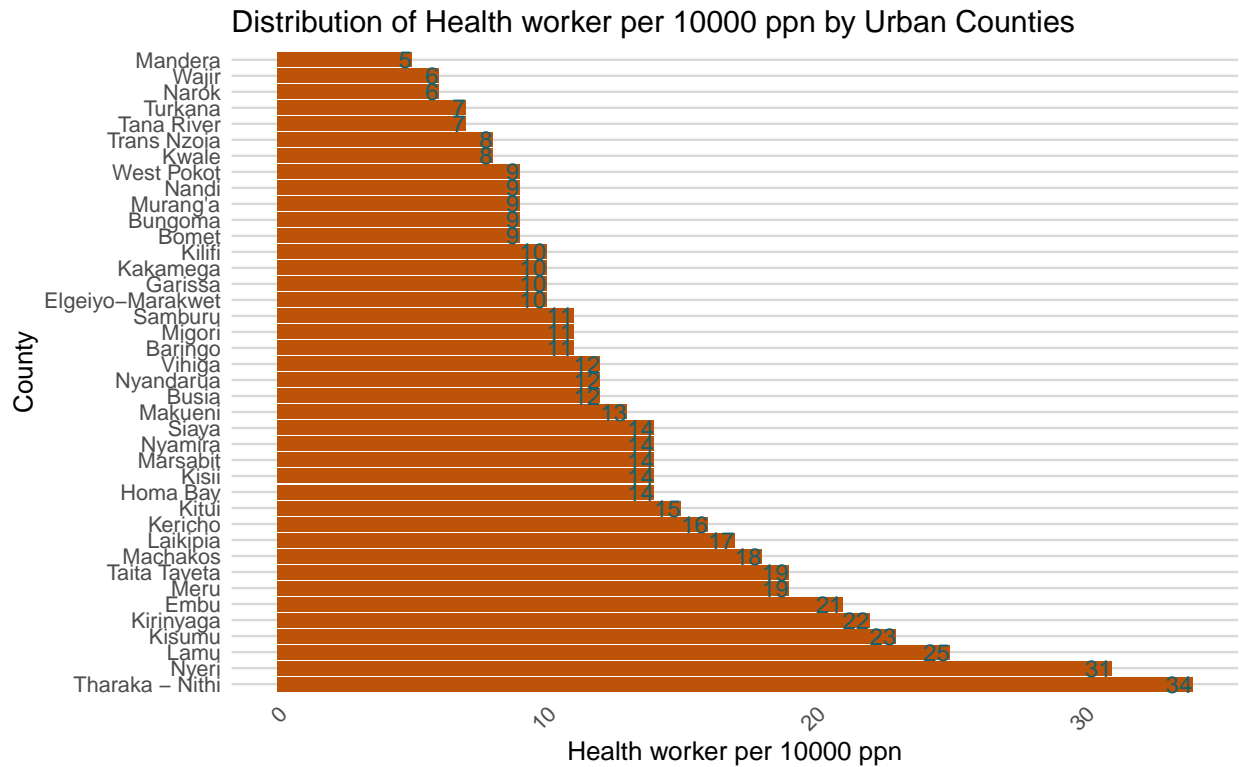
## Bar charts showing Hospital Bed density per 10000 population by counties

```
grid.arrange(fig7,fig8,ncol=1)
```



Barcharts showing Hospital worker per 10000 population Distribution by counties

```
grid.arrange(fig9,fig10,ncol=1)
```



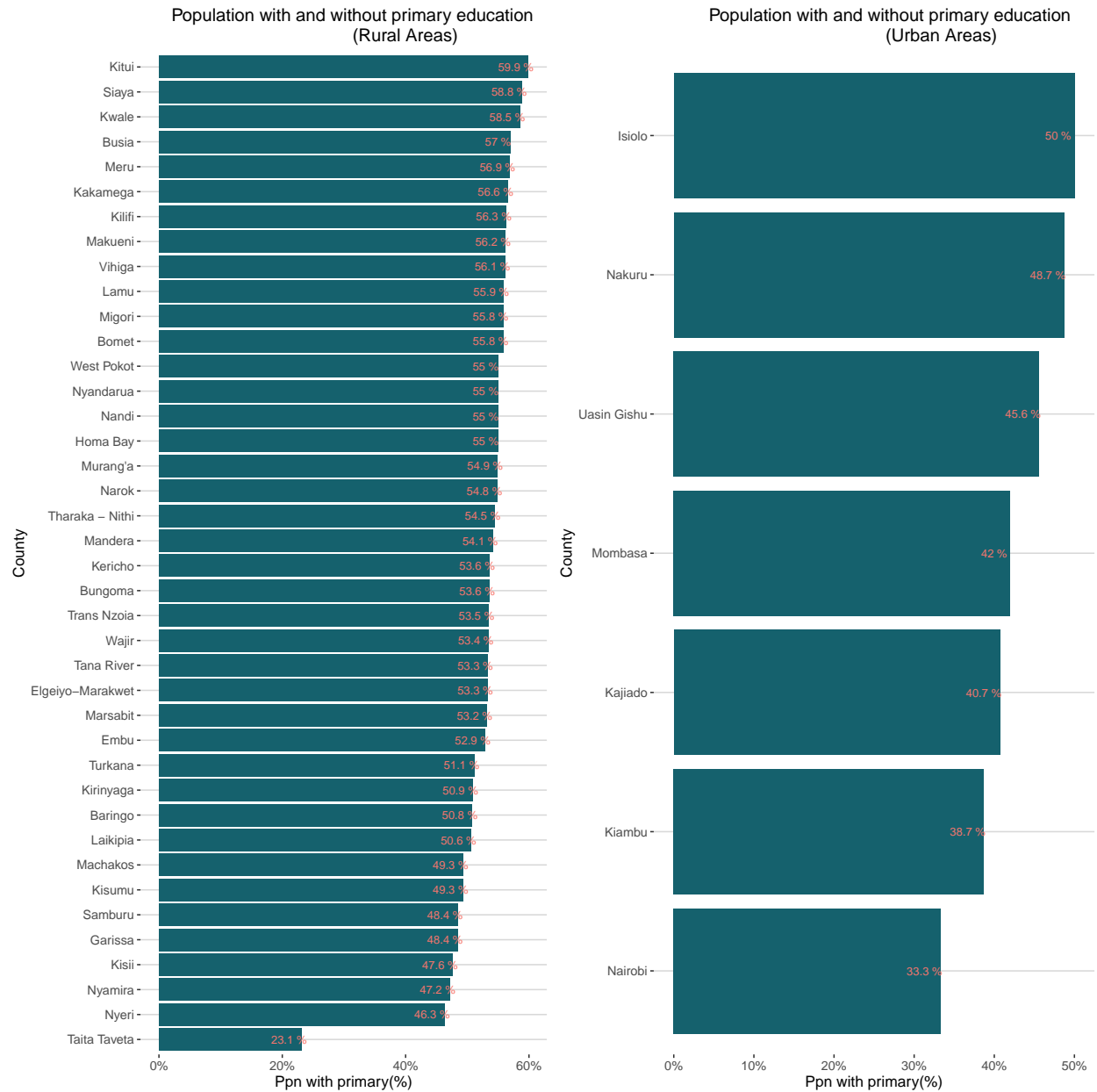
## Rural-Urban Education dynamics

## Primary school education

```
# find an arrangement index
y <- df %>% select(County, `Ppn with primary(%)` ,Status) %>% arrange(`Ppn with primary(%)` )
y$County <- factor(y$County, levels = y$County)

# create a graph for urban
rural <- y %>% filter(Status == 'Rural') %>% mutate(perc = paste(`Ppn with primary(%)` ,'%')) %>%
  ggplot(aes(x = County, y = `Ppn with primary(%)` ,label = perc)) + geom_bar(stat = 'identity', fill = colors[1]) +
  #scale_fill_manual(values = c(colors[1], colors[7])) +
  labs(title = "Population with and without primary education\n (Rural Areas)") + scale_y_continuous(label = perc) +
  geom_text(aes(y = `Ppn with primary(%)` - 2, color = colors[2]), size = 3, position = position_dodge(0.5)) +
  plot.title = element_text(hjust = 0.7) # Center the title
)+
  theme_hc()

# create a graph for urban
urban <- y %>% filter(Status == 'Urban') %>% mutate(perc = paste(`Ppn with primary(%)` ,'%')) %>%
  ggplot(aes(x = County, y = `Ppn with primary(%)` ,label = perc)) + geom_bar(stat = 'identity', fill = colors[1]) +
  #scale_fill_manual(values = c(colors[1], colors[7])) +
  labs(title = "Population with and without primary education\n (Urban Areas)") + scale_y_continuous(label = perc) +
  geom_text(aes(y = `Ppn with primary(%)` - 2, color = colors[2]), size = 3, position = position_dodge(0.5)) +
  plot.title = element_text(hjust = 0.7) # Center the title
)+
  theme_hc()
grid.arrange(rural, urban, ncol = 2)
```



## Secondary school education

```
# find an arrangement index
y <- df %>% select(County, `Ppn with secondary%`, Status) %>% arrange(`Ppn with secondary%`)
y$County <- factor(y$County, levels = y$County)

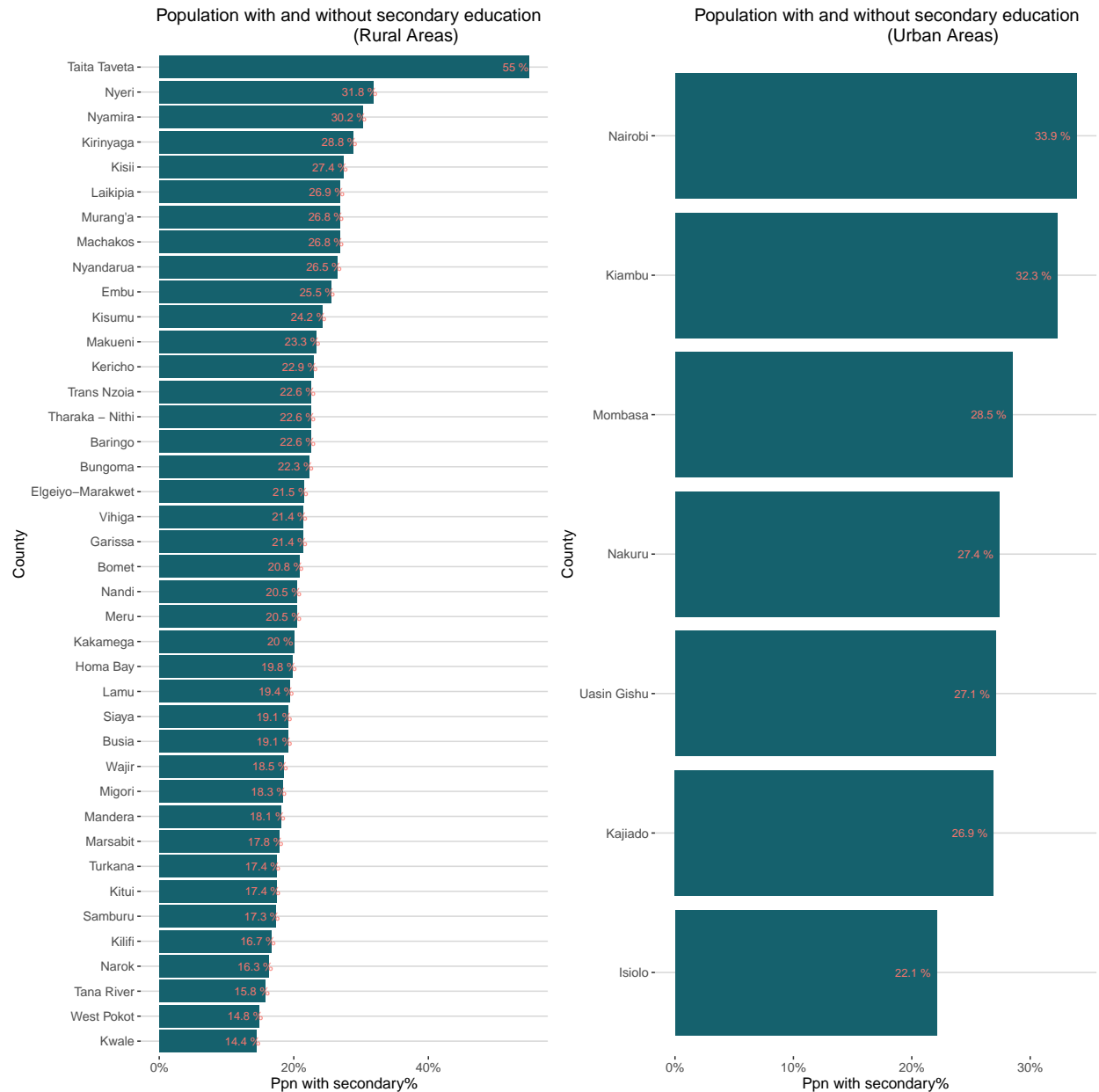
# create a graph for urban
rural <- y %>% filter(Status == 'Rural') %>% mutate(perc = paste(`Ppn with secondary%`, '%')) %>%
  ggplot(aes(x = County, y = `Ppn with secondary%`, label = perc)) + geom_bar(stat = 'identity', fill = color)
#scale_fill_manual(values = c(colors[1], colors[7])) +
```

```

labs(title = "Population with and without secondary education\n (Rural Areas)") + scale_y_continuous(la
geom_text(aes(y = `Ppn with secondary` - 2, color = colors[2]), size = 3, position = position_dodge(
  plot.title = element_text(hjust = 0.7) # Center the title
)+
theme_hc()

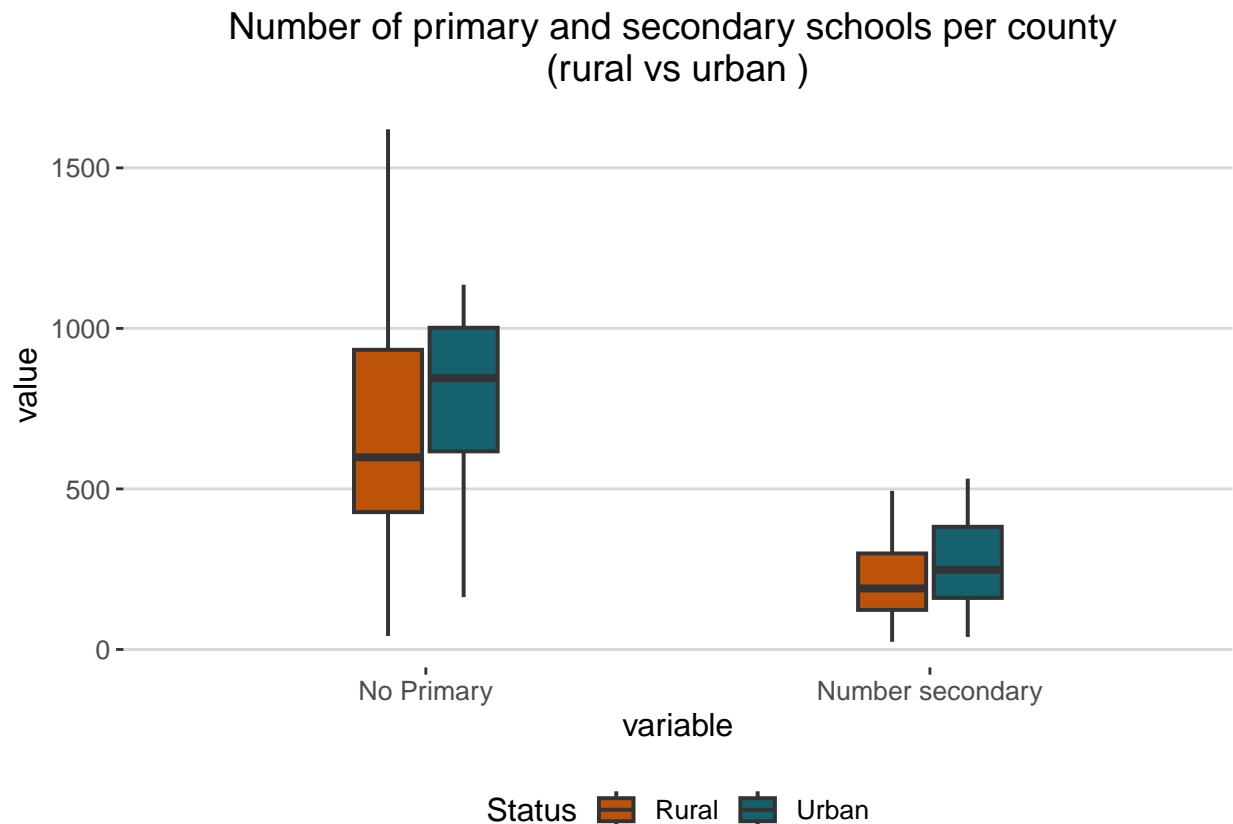
# create a graph for urban
urban <- y %>% filter(Status == 'Urban') %>% mutate(perc = paste(`Ppn with secondary`, '%')) %>%
ggplot(aes(x = County, y = `Ppn with secondary`, label = perc)) + geom_bar(stat = 'identity', fill = color
#scale_fill_manual(values = c(colors[1], colors[7])) +
labs(title = "Population with and without secondary education\n (Urban Areas)") + scale_y_continuous(la
geom_text(aes(y = `Ppn with secondary` - 2, color = colors[2]), size = 3, position = position_dodge(
  plot.title = element_text(hjust = 0.7) # Center the title
)+
theme_hc()
grid.arrange(rural, urban, ncol = 2)

```



```
df %>% select(Status, `No Primary`, `Number secondary`) %>%
  melt(id.vars = 'Status') %>%
  ggplot(aes(x = variable, y = value, fill = Status)) + geom_boxplot(width = 0.3, size = 0.7) +
  theme_hc() +
  labs(title = "Number of primary and secondary schools per county \n(rural vs urban)") +
  theme(
    plot.title = element_text(hjust = 0.5) # Center the title
  ) +
  scale_fill_manual(values = c(colors[1], colors[7]))
```





## Rural-Urban Misc

## Water and Sanitation

```
water_county <-kenyal %>%
  filter(Status== "Rural") %>%
  select("County","Status","Households with water %")
```

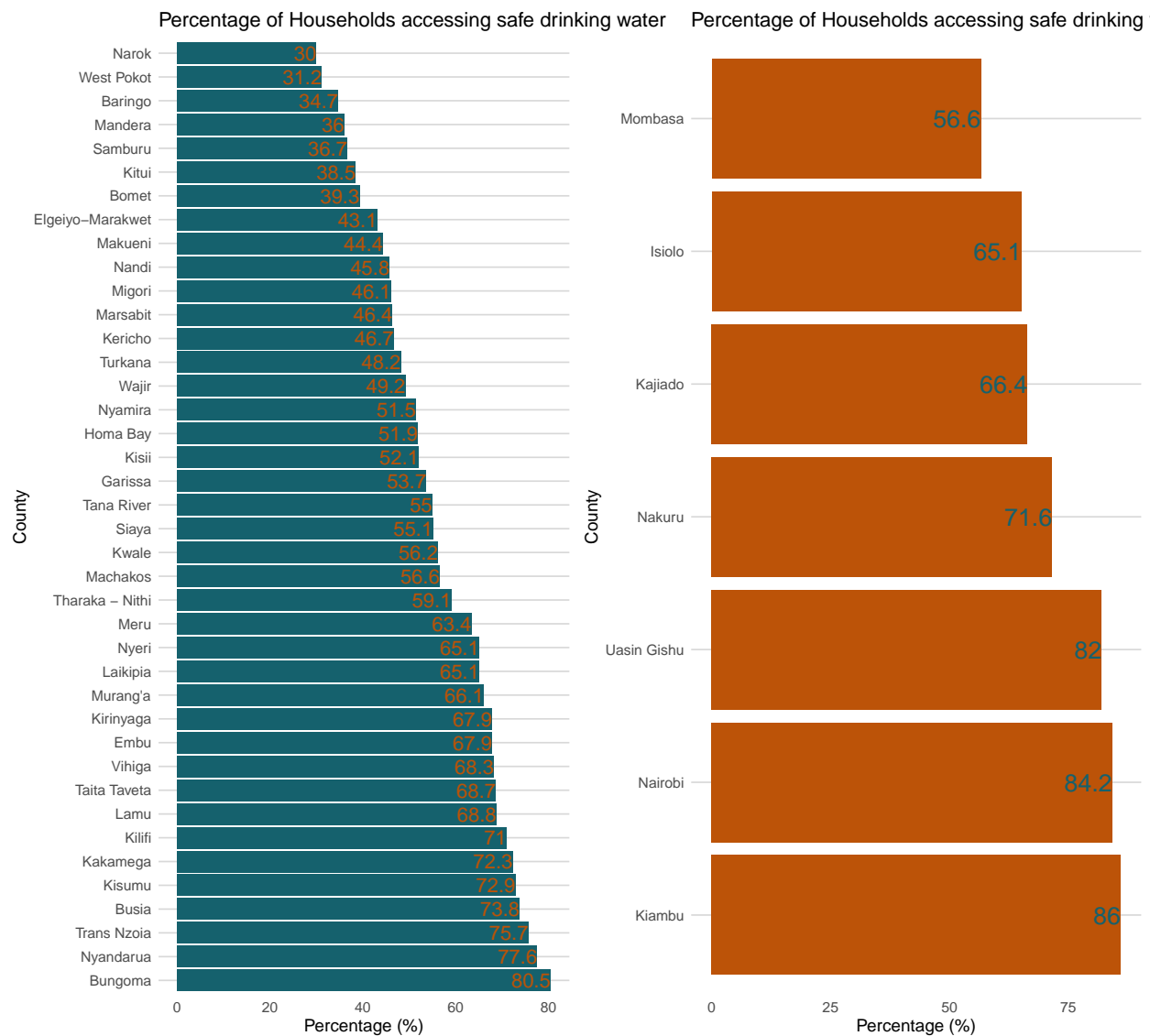
```
FIG1<- ggplot(water_county, aes(x = reorder(County, -`Households with water %`), y = `Households with water %`,
  geom_bar(stat = "identity", position = "dodge",show.legend = FALSE, fill = colors[7]) +
  geom_text(aes(label = `Households with water %`),color = colors[1], vjust = 0.5,hjust =1, size = 5))+
  labs(title = "Percentage of Households accessing safe drinking water",
    x = "County",
    y = "Percentage (%)") +
  theme_minimal()+
  coord_flip()+
  theme(panel.grid = element_blank())+
  theme_hc()
```

```
water_county <-kenyal %>%
  filter(Status== "Urban") %>%
  select("County","Status","Households with water %")
```

```
FIG2<- ggplot(water_county, aes(x = reorder(County, -`Households with water %`), y = `Households with water %`,
  geom_bar(stat = "identity", position = "dodge", show.legend = FALSE, fill = colors[1]) +
  geom_text(aes(label = `Households with water %`), vjust = 0.5, hjust = 1, size = 6, color = colors[7]) +
  labs(title = "Percentage of Households accessing safe drinking water",
    x = "County",
    y = "Percentage (%)") +
  theme_minimal() +
  coord_flip() +
  theme(panel.grid = element_blank()) + theme_hc()
```

Bargraphs showing the distribution of safe drinking water in the counties in terms of percentages.

```
grid.arrange(FIG1, FIG2, ncol=2)
```



Water plays a crucial role in the natural development of human beings. The following two graphs depict

the accessibility of water in various counties. Among the rural counties, Bungoma leads with an impressive accessibility rate of 80.5%. Conversely, among the urban counties, Kiambu boasts a remarkable 86% accessibility rate, surpassing Nairobi, the capital city of Kenya. These findings underscore the significance of water accessibility in both rural and urban settings.

```
# water_county
```

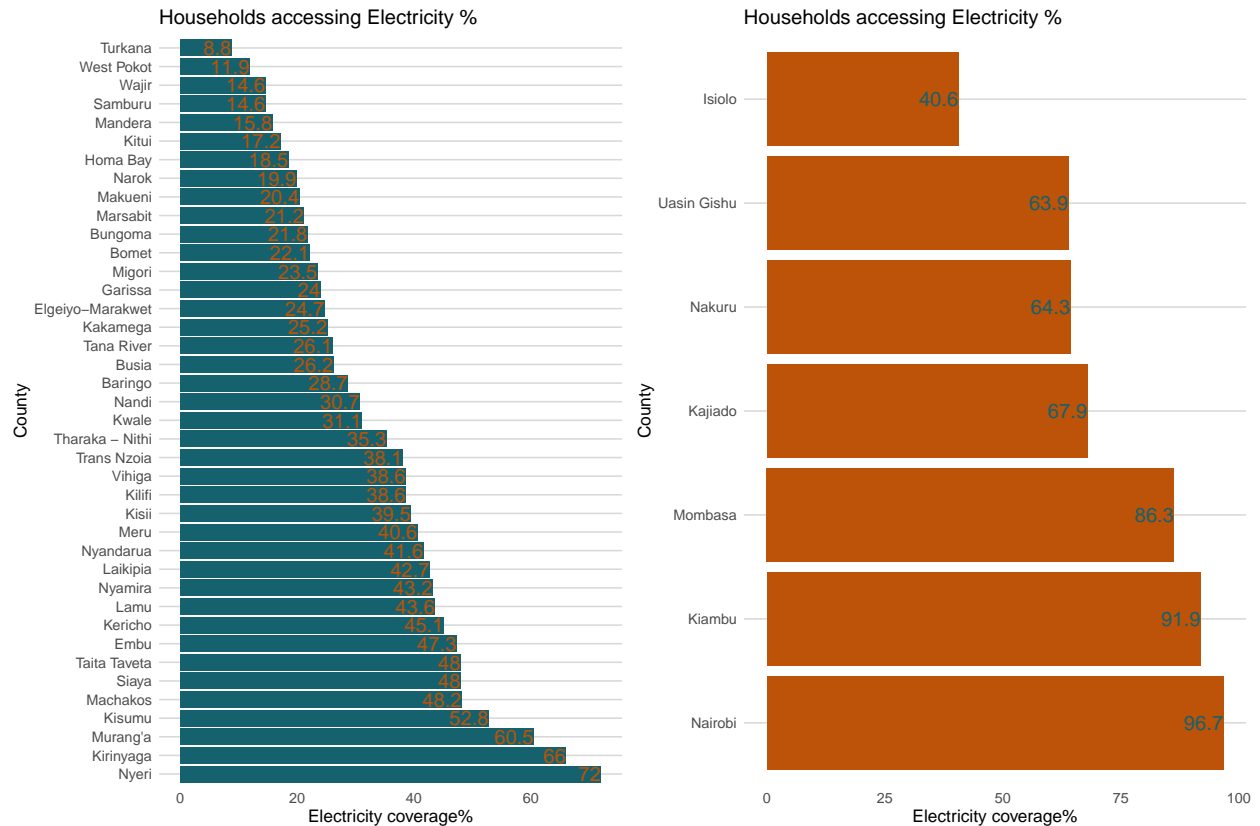
## Energy

```
Electricity_county<-kenya1 %>%
  filter(Status=="Rural") %>%
  select("County","Status","Household with Electricity %") %>%
  arrange(`Household with Electricity %`)
FIG3<-ggplot(Electricity_county, aes(x = reorder(County, -`Household with Electricity %`), y = `Household with Electricity %`)) +
  geom_bar(stat = "identity", position = "dodge",show.legend = FALSE, fill= colors[7]) +
  geom_text(aes(label = `Household with Electricity %`), vjust = 0.5,hjust =1, size = 5,color= colors[1]) +
  labs(title = "Households accessing Electricity %",
       x = "County",
       y = "Electricity coverage%") +
  theme_minimal()+
  coord_flip()+
  theme(panel.grid = element_blank())+theme_hc()
```

```
Electricity_county<-kenya1 %>%
  filter(Status=="Urban") %>%
  select("County","Status","Household with Electricity %") %>%
  arrange(`Household with Electricity %`)
FIG4<-ggplot(Electricity_county, aes(x = reorder(County, -`Household with Electricity %`), y = `Household with Electricity %`)) +
  geom_bar(stat = "identity", position = "dodge",show.legend = FALSE,fill= colors[1]) +
  geom_text(aes(label = `Household with Electricity %`), vjust = 0.5,hjust =1, size = 5,color= colors[7]) +
  labs(title = "Households accessing Electricity %",
       x = "County",
       y = "Electricity coverage%") +
  theme_minimal()+
  coord_flip()+
  theme(panel.grid = element_blank())+
  theme_hc()
```

## Bargraphs showing the energy distribution in the Counties in percentages

```
grid.arrange(FIG3,FIG4,ncol=2)
```



The level of energy connectivity is a pivotal indicator of societal development. The two aforementioned graphs provide insights into the extent of connectivity across various counties. Within the rural category, Turkana County registers the lowest electricity connection rate at 8.8%, while Nyeri County leads with an impressive 72% of households connected to the grid. Among urban counties, Isiolo County reports a 40.6% access to energy, while the capital city boasts an impressive 96.7% electricity accessibility rate. These findings underscore the importance of energy connectivity as a key driver of development in both rural and urban areas.

##Rural Access Index (RAI)-%

```
Rural_index_county<-kenya1 %>%
  filter(Status=="Rural") %>%
  select("County","Status","Rural Access Index%")
```

```
FIG5<- ggplot(Rural_index_county, aes(x = reorder(County,`Rural Access Index%`), y = `Rural Access Index%`)) +
  geom_bar(stat = "identity", position = "dodge",show.legend = FALSE,fill = colors[7]) +
  geom_text(aes(label = `Rural Access Index%`), vjust = 0.5,hjust =1, size = 5,color = colors[1])+
  labs(title = "Rural access Index (RAI) % ~ Rural Counties",
       x = "County",
       y = "percentage (%)") +
  theme_minimal()+
  scale_fill_manual(values = colors)+
  coord_flip()+
  theme(panel.grid = element_blank())+
  theme_hc()
```

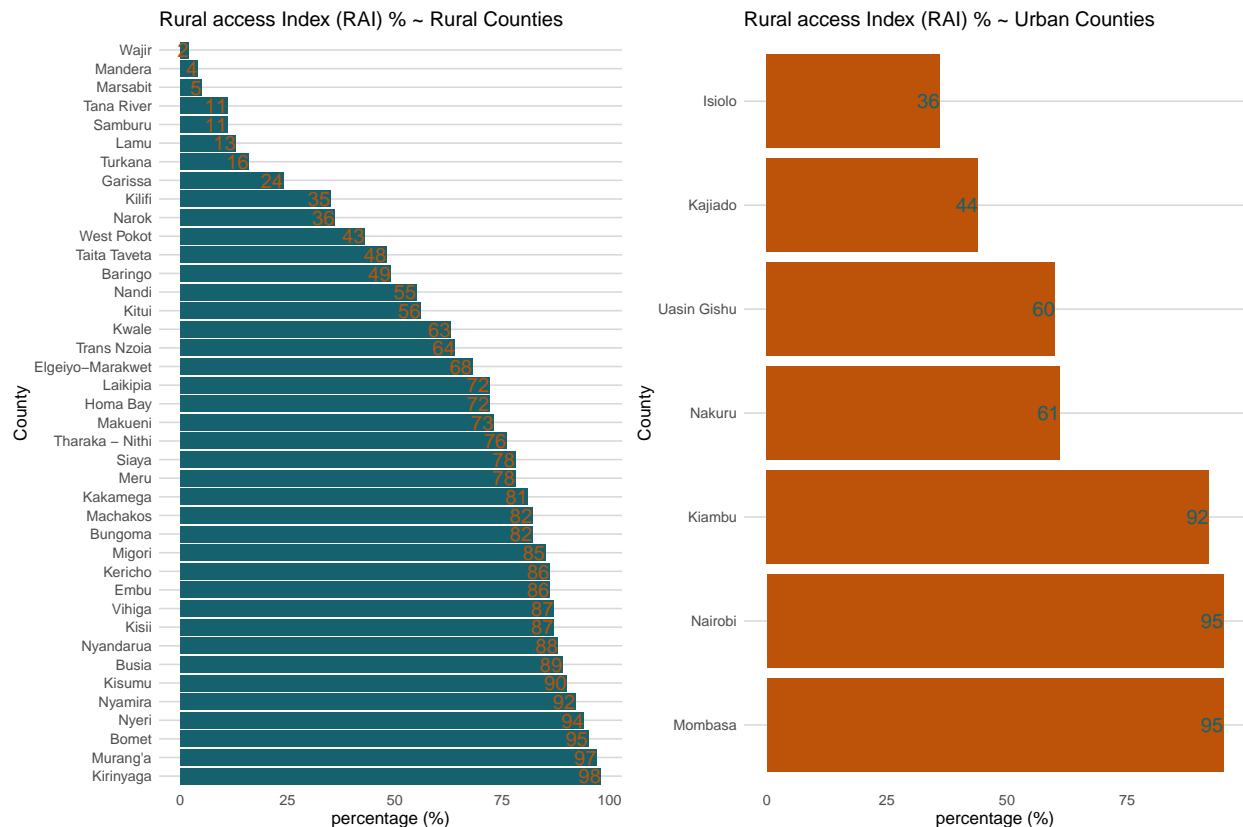
```
Rural_index_county<-kenya1 %>%
  filter(Status=="Urban") %>%
```

```
select("County","Status","Rural Access Index%")
```

```
FIG6<- ggplot(Rural_index_county, aes(x = reorder(County,`Rural Access Index`), y = `Rural Access Index`)) +
  geom_bar(stat = "identity", position = "dodge", show.legend = FALSE, fill = colors[1]) +
  geom_text(aes(label = `Rural Access Index`), vjust = 0.5, hjust = 1, size = 5, color = colors[7]) +
  labs(title = "Rural access Index (RAI) % ~ Urban Counties ",
        x = "County",
        y = "percentage (%)") +
  theme_minimal() +
  scale_fill_manual(values = colors) +
  coord_flip() +
  theme(panel.grid = element_blank()) +
  theme_hc()
```

##Bargraphs showing Rural Access Index (RAI)-%

```
grid.arrange(FIG5, FIG6, ncol=2)
```



Robust infrastructure is a cornerstone of thriving communities, facilitating seamless mobility for both goods and people. Quality infrastructure plays a pivotal role in fostering regional development. The two accompanying graphs present the Rural Access Index in terms of percentages. It is particularly noteworthy that Kirinyaga and Murang'a, despite being categorized as rural counties, exhibit superior rural accessibility compared to Nairobi and Mombasa, typically classified as urban counties. This observation underscores the significance of infrastructure in rural areas and challenges conventional urban-rural distinctions.

##Information Communication and Technology (ICT)

```
Internet_county<-kenya1 %>%
  filter(Status == "Rural") %>%
  select("County", "Status", "population accessing internet%")
```

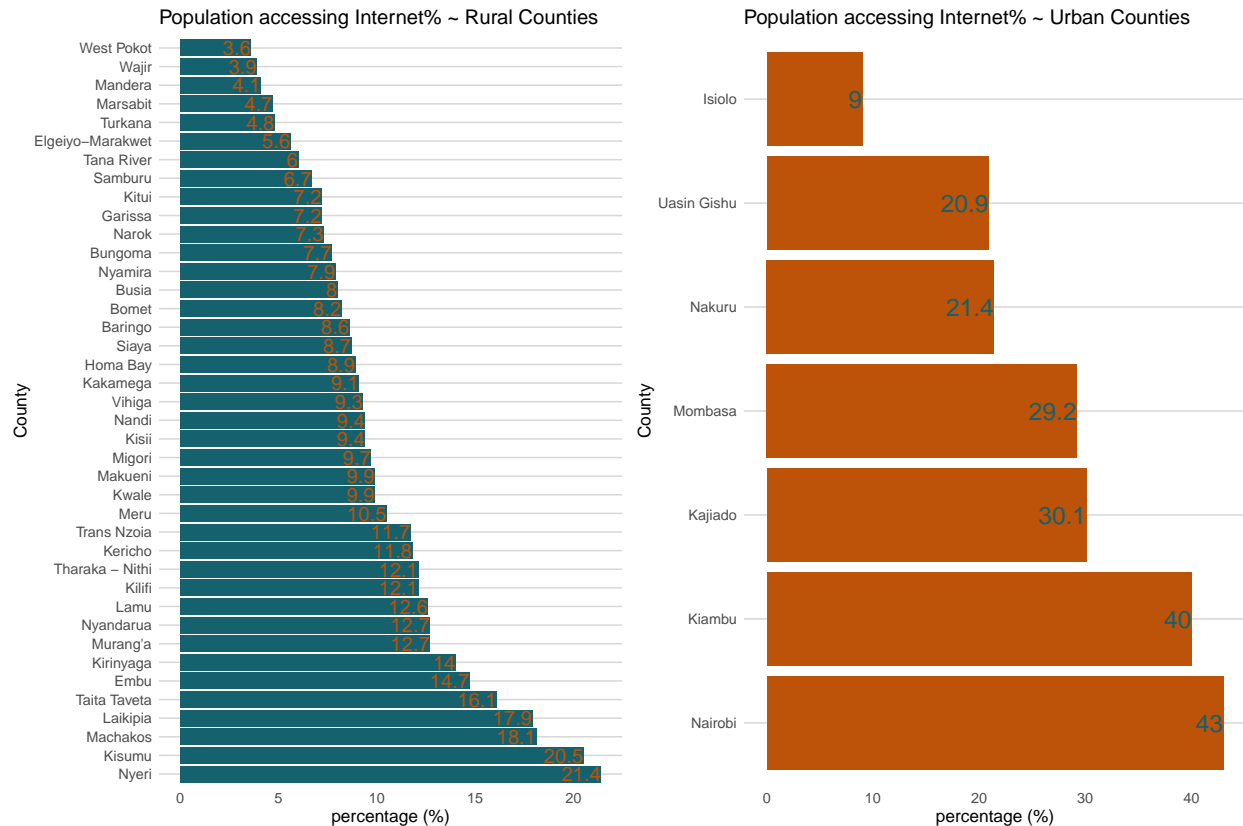
```
FIG7<- ggplot(Internet_county, aes(x = reorder(County, `population accessing internet%`), y = `population accessing internet%`)) +
  geom_bar(stat = "identity", position = "dodge", show.legend = FALSE, fill = colors[7]) +
  geom_text(aes(label = `population accessing internet%`), vjust = 0.5, hjust = 1, size = 5, color = colors[7]) +
  labs(title = "Population accessing Internet% ~ Rural Counties",
       x = "County",
       y = "percentage (%)") +
  theme_minimal() +
  scale_fill_manual(values = colors) +
  coord_flip() +
  theme(panel.grid = element_blank()) +
  theme_hc()
```

```
Internet_county<-kenya1 %>%
  filter(Status == "Urban") %>%
  select("County", "Status", "population accessing internet%")
```

```
FIG8<-ggplot(Internet_county, aes(x = reorder(County, `population accessing internet%`), y = `population accessing internet%`)) +
  geom_bar(stat = "identity", position = "dodge", show.legend = FALSE, fill = colors[1]) +
  geom_text(aes(label = `population accessing internet%`), vjust = 0.5, hjust = 1, size = 6, color = colors[1]) +
  labs(title = "Population accessing Internet% ~ Urban Counties",
       x = "County",
       y = "percentage (%)") +
  theme_minimal() +
  scale_fill_manual(values = colors) +
  coord_flip() +
  theme(panel.grid = element_blank()) +
  theme_hc()
```

##Bargraphs showing accessibility to Internet (%)

```
grid.arrange(FIG7, FIG8, ncol=2)
```



Accessibility to the internet is a fundamental driver of the digital economy. The two provided graphs offer insights into the percentage of the population with internet access. Notably, West Pokot County has the lowest internet accessibility, standing at a mere 3.6%. In contrast, Nyeri County leads among rural counties with a 21.4% internet penetration rate.

In the urban category, the capital city takes the lead with an impressive 43% internet accessibility, followed closely by Kiambu County at 40%. It's noteworthy that rural counties generally exhibit lower internet accessibility rates. These findings underscore the importance of bridging the digital divide to promote equitable access to the internet, which is essential for fostering the digital economy.

## ##Mobile Phone Ownership

```
mobile_phones<-kenya1 %>%
  filter(Status == "Rural") %>%
  select("County","Status","population owning mobile phones%")
```

```
FIG9<- ggplot(mobile_phones, aes(x = reorder(County, -`population owning mobile phones%`), y = `population owning mobile phones%`)) +
  geom_bar(stat = "identity", position = "dodge", show.legend = FALSE, fill = colors[7]) +
  geom_text(aes(label = `population owning mobile phones%`), vjust = 0.5, hjust = 1, size = 5, color = colors[7]) +
  labs(title = "Population owning mobile phones% ~ Rural Counties",
       x = "County",
       y = "percentage (%)") +
  theme_minimal()+
  scale_fill_manual(values = colors)+
  coord_flip()+theme(panel.grid = element_blank())
```

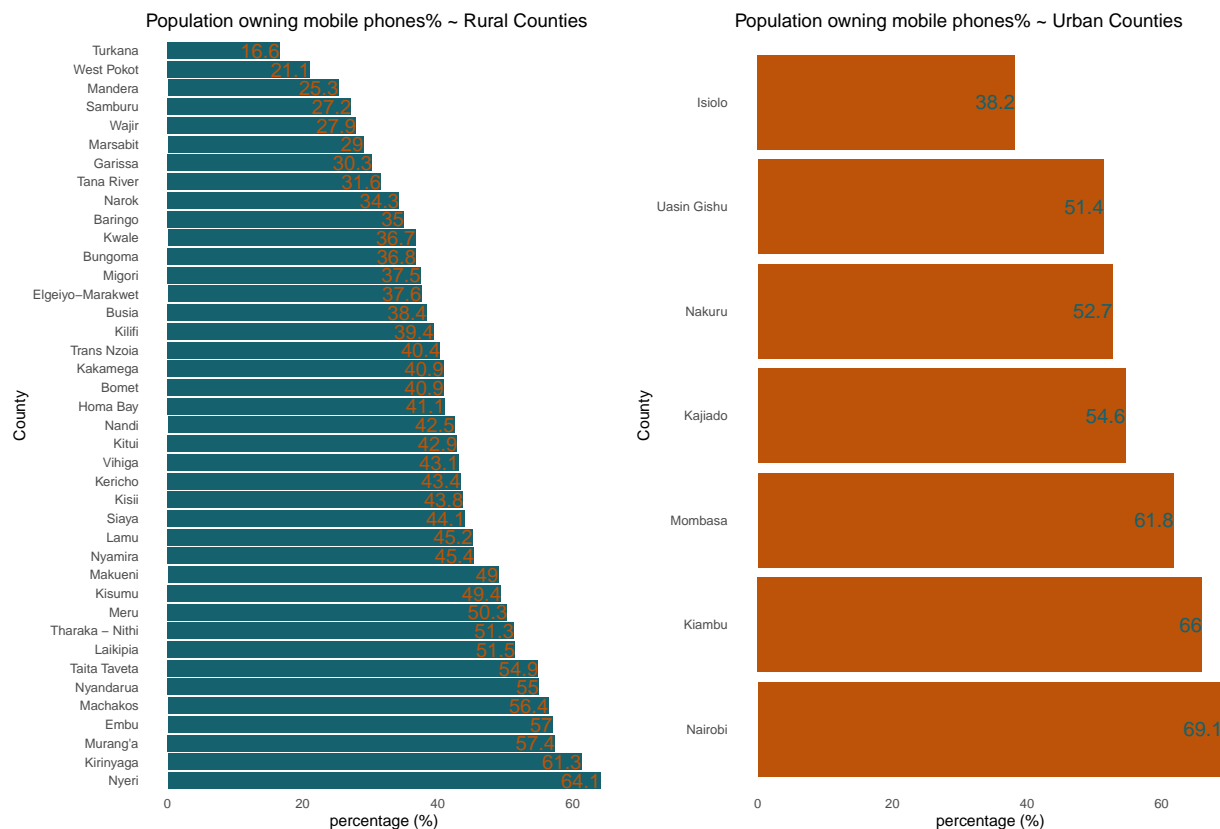
```
mobile_phones<-kenya1 %>%
  filter(Status == "Urban") %>%
```

```
select("County","Status","population owning mobile phones%")
```

```
FIG10<- ggplot(mobile_phones, aes(x = reorder(County, -`population owning mobile phones%`), y = `population owning mobile phones%`)) +
  geom_bar(stat = "identity", position = "dodge", show.legend = FALSE, fill = colors[1]) +
  geom_text(aes(label = `population owning mobile phones%`), vjust = 0.5, hjust = 1, size = 5, color = colors[2]) +
  labs(title = "Population owning mobile phones% ~ Urban Counties",
       x = "County",
       y = "percentage (%)") +
  theme_minimal() +
  scale_fill_manual(values = colors) +
  coord_flip() + theme(panel.grid = element_blank())
```

## Bargraphs showing ownership of mobile phones(%)

```
library(gridExtra)
grid.arrange(FIG9, FIG10, ncol=2)
```



Mobile phone products and services have played a pivotal role in driving the economic development of communities. The provided graphs highlight significant disparities in mobile phone ownership across counties. Turkana County reports a relatively low mobile phone ownership rate, with only 16.6% of the population owning a mobile device. In sharp contrast, Nairobi leads the way with a substantial 69.1% of its population owning a mobile phone.

These disparities in mobile phone ownership underscore the digital divide between regions and highlight the potential for mobile technology to contribute to economic growth and connectivity, especially in areas with lower adoption rates.



## findings and conclusion

In conclusion, the findings from the data presented reveal the critical role of infrastructure and technology in shaping the development landscape of communities. Access to essential services such as water, electricity, and the internet, along with mobile phone ownership, are pivotal factors influencing the economic well-being of regions. Notably, these findings challenge traditional urban-rural distinctions, as certain rural counties exhibit superior infrastructure and technology accessibility compared to urban counterparts. Bridging the digital and infrastructural divides is essential for fostering equitable development and ensuring that the benefits of the digital economy are accessible to all, regardless of geographic location. These insights underscore the need for targeted policies and investments to promote inclusive growth and sustainable development across diverse communities.