An Analysis of BRT Systems across the World.

Jeffrey Ohene

2022-08-28

Figure 1: Bus Rapid Transit, Curitiba

The BRT system, short for Bus Rapid Transit system, according to Wikipedia, is any bus system with designated & dedicated roadways for the buses. Priority is given to these buses over any other vehicle at intersections and traffic lights. The system offers faster speed, dependability at lower costs than the traditional bus system while efficiently making far better capacity design. This project is going to analyse the BRT systems of Africa, Asia, Europe and North-America and then Latin America.

The analysis will largely be based on the key indicators such as passengers per day, number of corridors and length of corridors of each continent and visualize how they stack up against the other continents.

```
#Load libraries for analysis

library(readr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
## Warning in as.POSIXlt.POSIXct(Sys.time()): unable to identify current timezone 'C':
## please set environment variable 'TZ'
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --

## v tibble  3.1.7     v stringr 1.4.0
## v tidyr   1.2.0     v forcats 0.5.1
## v purrr   0.3.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(scales)
```

```
##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
##
## The following object is masked from 'package:readr':
##
##     col_factor
```

We would begin with Asia. Asia according to the Global BRT Data Organization features 12 countries which use the BRT system. For the sake of data integrity, Malaysia, which at the time of analysis had missing information for passengers per day was omitted. A number of cities with missing data for the same indicator were also omitted and all omitted countries and cities will be listed in the appendix.

First of, the data for the 11 Asian countries with a BRT will be loaded below and joined and stored under Asia, so we can have all the cities of these countries in a single place for further analysis.

```
#Load datasets for Asian countries

china <- read_csv("brtdata-china.csv")
```

```
## Rows: 18 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
india <- read_csv("brtdata-india.csv")
```

```
## Rows: 9 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
indonesia <- read_csv("brtdata-indonesia.csv")
```

```
## Rows: 1 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
iran <- read_csv("brtdata-iran.csv")
```

```
## Rows: 2 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
israel <- read_csv("brtdata-israel.csv")
```

```
## Rows: 1 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
japan <- read_csv("brtdata-japan.csv")
```

```
## Rows: 1 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
pakistan <- read_csv("brtdata-pakistan.csv")
```

```
## Rows: 2 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
rep_of_korea <- read_csv("brtdata-republic_of_korea.csv")
```

```
## Rows: 1 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
taiwan <- read_csv("brtdata-taiwan.csv")
```

```
## Rows: 3 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
thailand <- read_csv("brtdata-thailand.csv")
```

```
## Rows: 1 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
```

```
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
vietnam <- read_csv("brtdata-vietnam.csv")
```

```
## Rows: 1 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
#Bind datasets
asien <- rbind(
  china,india,indonesia,iran,israel,japan,pakistan,
  rep_of_korea,taiwan,thailand,vietnam)

asien
```

```
## # A tibble: 40 x 4
##    Cities    `Passengers per Day` `Number of Corridors` `Length (km)`
##    <chr>                    <dbl>                 <dbl>         <dbl>
##  1 Beijing                 305000                     4            75
##  2 Changzhou               350000                     2            50
##  3 Chongqing                12000                     1            12
##  4 Dalian                   87000                     1            14
##  5 Guangzhou               850000                     1            23
##  6 Hangzhou                260000                     3            55
##  7 Hefei                    65250                     1             8
##  8 Jinan                   220000                     6            61
##  9 Kunming                 156000                     1            47
## 10 Lanzhou                 290000                     1             9
## # ... with 30 more rows
```

Voila! Now that we have all our Asian cities in a single place, we can proceed to rename the columns in a way that will make our analysis smooth as some of the column strings have words R might interpret as commands.

```
#Rename columns for more efficient identification

asien <- asien %>%
  rename(
    city=Cities,
    passengers=`Passengers per Day`,
    corridors=`Number of Corridors`,
    length=`Length (km)`
  )

asien
```

```
## # A tibble: 40 x 4
##    city      passengers corridors length
##    <chr>          <dbl>     <dbl>  <dbl>
##  1 Beijing       305000         4     75
##  2 Changzhou     350000         2     50
##  3 Chongqing      12000         1     12
##  4 Dalian         87000         1     14
##  5 Guangzhou     850000         1     23
##  6 Hangzhou      260000         3     55
##  7 Hefei          65250         1      8
##  8 Jinan         220000         6     61
##  9 Kunming       156000         1     47
## 10 Lanzhou       290000         1      9
## # ... with 30 more rows
```

```
str(asien)
```

```
## spec_tbl_df [40 x 4] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ city      : chr [1:40] "Beijing" "Changzhou" "Chongqing" "Dalian" ...
##  $ passengers: num [1:40] 305000 350000 12000 87000 850000 ...
##  $ corridors : num [1:40] 4 2 1 1 1 3 1 6 1 1 ...
##  $ length    : num [1:40] 75 50 12 14 23 55 8 61 47 9 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Cities = col_character(),
##   ..   `Passengers per Day` = col_number(),
##   ..   `Number of Corridors` = col_double(),
##   ..   `Length (km)` = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

Smooth! We might be curious. What is the total length of corridors in Asia? or maybe we want to know, how many passengers use the BRT system per Day in Asia. Would that be a huge number? Or maybe Asians do not really like the BRT system? Enough speculation, time for facts!

```
#Calculate daily passenger total & Corridor length

tot_asi <- asien %>%
  summarize(
    passengers_per_day=sum(passengers),
    total_distance=sum(length)
  )

tot_asi
```

```
## # A tibble: 1 x 2
##   passengers_per_day total_distance
##                <dbl>          <dbl>
## 1            9238060           1602
```

As we can see, some 9.2 million people use the BRT system every single day in Asia. The total length of BRT corridors is also 1602 km long!

Now we know a large number of people use the BRT in some form in Asia on a daily basis. Wouldn't it be more interesting to know which cities rank top? If anyone had a wild guess, they would probably suggest a city in China or India as both of them are the most populated countries in the world. Nice try, but let's see what our data says!

```
#Top 10 cities by daily passenger count

top_cities_asien <- asien %>%
  select(city,passengers,corridors,length) %>%
  arrange(desc(passengers))%>%
  slice(1:3)

top_cities_asien
```

```
## # A tibble: 3 x 4
##   city      passengers corridors length
##   <chr>          <dbl>     <dbl>  <dbl>
## 1 Tehran       2000000         8    130
## 2 Taipei       1302832         1     60
## 3 Guangzhou     850000         1     23
```

Figure 2: Tehran

Interestingly, Tehran, Iran, has the highest BRT system patronage out of all Asian cities. They also have the longest corridor length and second longest number of corridors! Impressive.

Taiwan's beautiful Taipei comes in second and has 1.3 million daily passengers. This is all the more impressive when you consider the fact that the city of Taipei has only a single corridor, ranking it joint bottom among all the Asian cities and also they have the 5th longest corridor length.

Only Tehran and Taipei have over a million daily passengers in Asia. China's Guangzhou comes in third with a daily passenger total of 850,000, a single corridor which is only 23km.

Now to find some even more insights, let us manipulate our data further.

```
#Calculate mean of the passenger distribution
avg_asi <- asien %>%
  select(city,passengers) %>%
  summarize(
    avg=mean(passengers),
    avg_km=mean(asien$length)
  )

avg_asi
```

```
## # A tibble: 1 x 2
##       avg avg_km
##     <dbl>  <dbl>
## 1 230952.   40.0
```

Averagely, the number daily passengers is approximately 231,000 and the total distance of the corridors on the continent is about 40km.

```
#Split to quartiles to check real structure of passenger distribution
quantile(asien$passengers)
```

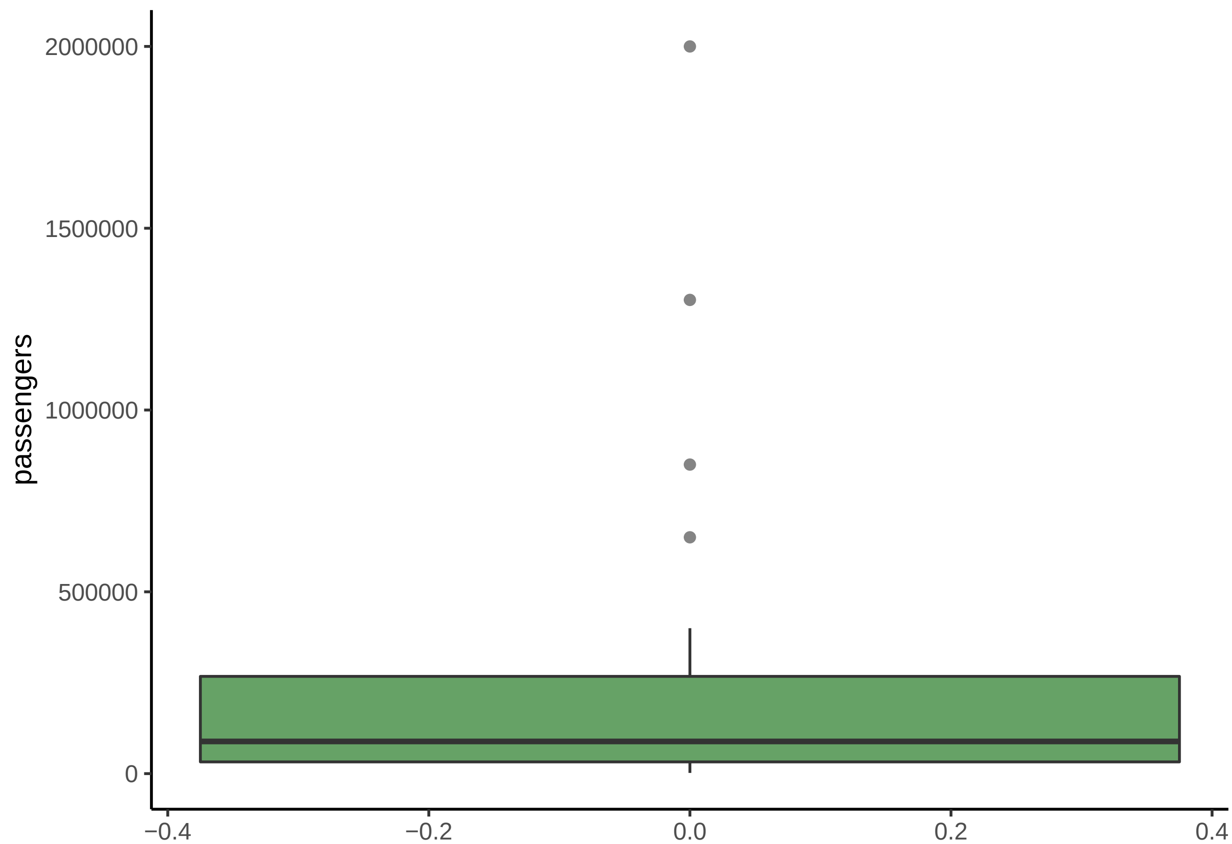
```
##      0%     25%     50%     75%    100%
##    2000   32250   88500  267500 2000000
```

```
IQR(asien$passengers)
```

```
## [1] 235250
```

```
#Visualize quartiles
p0 <- ggplot(asien,aes(passengers)) +
  geom_boxplot(fill="darkgreen", alpha=0.6)

p0 + theme_classic() +
  coord_flip()
```



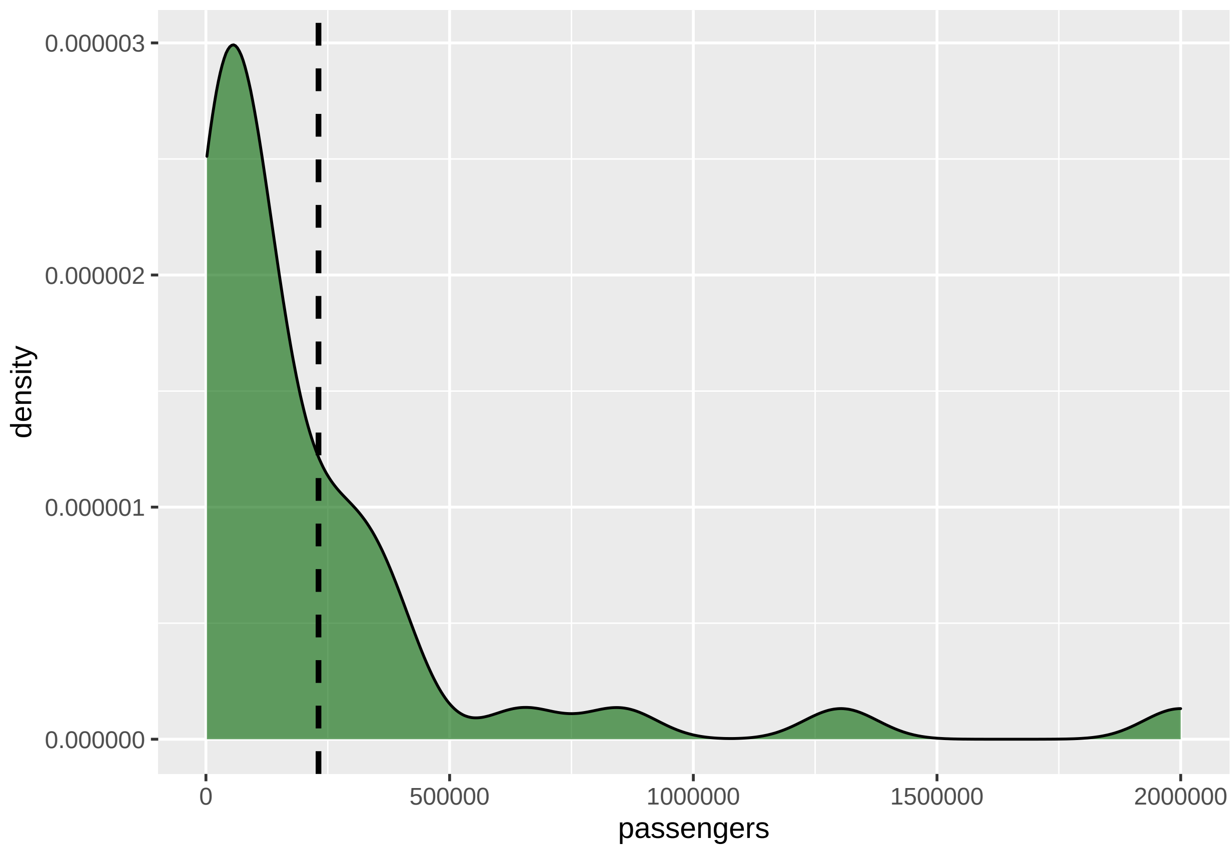
Quartiles help us measure the spread of our data. Here we find out that Tehran and Taipei, are the farthest outliers and our data is structured more around the 32,250 - 267,500 range. We can conclude that, realistically, Asian cities' patronage of the BRT system is around the 235,250.

Our density plot below helps us confirm the above statistic that more Asian cities have a daily passenger total of below approximately 235,250. The plot also shows that any given city chosen at random is likely to have a daily passenger number of below 125,000.

```
#Density plot for Asia
set.seed(250)

p1 <- ggplot(asien,aes(passengers)) +
  geom_density(fill="darkgreen", alpha=0.6)

p1 + geom_vline(aes(xintercept=mean(asien$passengers)),
            color="black", linetype="dashed", size=1)+
  scale_y_continuous(labels=comma)
```



The next continent that would be subject to our analysis will be Africa. Let us clean and model our data nicely for further analysis.

```
#Load datasets for African countries
nigeria <- read_csv("brtdata-nigeria.csv")
```

```
## Rows: 1 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
south_afrika <- read_csv("brtdata-south_africa.csv")
```

```
## Rows: 3 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
tanzania <- read_csv("brtdata-tanzania.csv")
```

```
## Rows: 1 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
#Bind datasets into a single dataframe
afrika <- rbind(
  nigeria,south_afrika,tanzania
)

afrika
```

```
## # A tibble: 5 x 4
##   Cities        `Passengers per Day` `Number of Corridors` `Length (km)`
##   <chr>                        <dbl>                 <dbl>         <dbl>
## 1 Lagos                       200000                     1            22
## 2 Cape Town                    66178                     2            31
## 3 Johannesburg                 42000                     2            44
## 4 Pretoria                      3400                     2            14
## 5 Dar es Salaam               180000                     1            21
```

```r
#Rename columns for more efficient identification
afrika <- afrika %>%
  rename(
    city=Cities,
    passengers=`Passengers per Day`,
    corridors=`Number of Corridors`,
    length=`Length (km)`
  )

afrika
```

```
## # A tibble: 5 x 4
##   city          passengers corridors length
##   <chr>              <dbl>     <dbl>  <dbl>
## 1 Lagos             200000         1     22
## 2 Cape Town          66178         2     31
## 3 Johannesburg       42000         2     44
## 4 Pretoria            3400         2     14
## 5 Dar es Salaam     180000         1     21
```

```r
str(afrika)
```

```
## spec_tbl_df [5 x 4] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ city      : chr [1:5] "Lagos" "Cape Town" "Johannesburg" "Pretoria" ...
```

```
##  $ passengers: num [1:5] 200000 66178 42000 3400 180000
##  $ corridors : num [1:5] 1 2 2 2 1
##  $ length    : num [1:5] 22 31 44 14 21
##  - attr(*, "spec")=
##   .. cols(
##   ..   Cities = col_character(),
##   ..   `Passengers per Day` = col_number(),
##   ..   `Number of Corridors` = col_double(),
##   ..   `Length (km)` = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
#Sum of daily passengers and corridor distance
tot_afr <- afrika %>%
  summarize(
    passengers_per_day=sum(passengers),
    total_distance=sum(length)
  )

tot_afr
```

```
## # A tibble: 1 x 2
##   passengers_per_day total_distance
##                <dbl>          <dbl>
## 1             491578            132
```

A total number of 491,578 people use the BRT system on a daily basis in Africa. The total length of BRT corridors in this vast continent is disappointingly only 132km long!

```
#Top 3 African cities by daily passenger total
top_cities_afrika <- afrika %>%
  select(city,passengers,corridors,length) %>%
  arrange(desc(passengers))%>%
  slice(1:3)

top_cities_afrika
```

```
## # A tibble: 3 x 4
##   city          passengers corridors length
##   <chr>              <dbl>     <dbl>  <dbl>
## 1 Lagos             200000         1     22
## 2 Dar es Salaam     180000         1     21
## 3 Cape Town          66178         2     31
```

Nigeria's Lagos has the most daily passenger total with 200,000 people. There is only a single corridor in the city which spans over 22km. Dar es Salaam in Tanzania comes in close second with 180,000 daily passengers using the BRT system in a single corridor of length 21km. Cape Town in South Africa, with surprisingly the more corridors than Lagos and Dar es Salaam that span 31km(more than each of the top two) have an underwhelming amount of daily passengers with only around 66,000 daily passengers.

```
#Calculate mean of the passenger distribution
avg_afr <- afrika %>%
```

Figure 3: Lagos

```
  select(city,passengers) %>%
  summarize(
    avg_ppl=mean(passengers),
    avg_km=mean(afrika$length)
  )

avg_afr
```

```
## # A tibble: 1 x 2
##   avg_ppl avg_km
##     <dbl>  <dbl>
## 1  98316.   26.4
```

On an average, the continent of Africa have approximately 98,000 daily passengers This figure is over two times less than the average daily passengers in Asia. The average length of a given corridor in a city on the continent is 26.4km

```
#Split to quartiles to check real structure of passenger distribution
quantile(afrika$passengers)
```

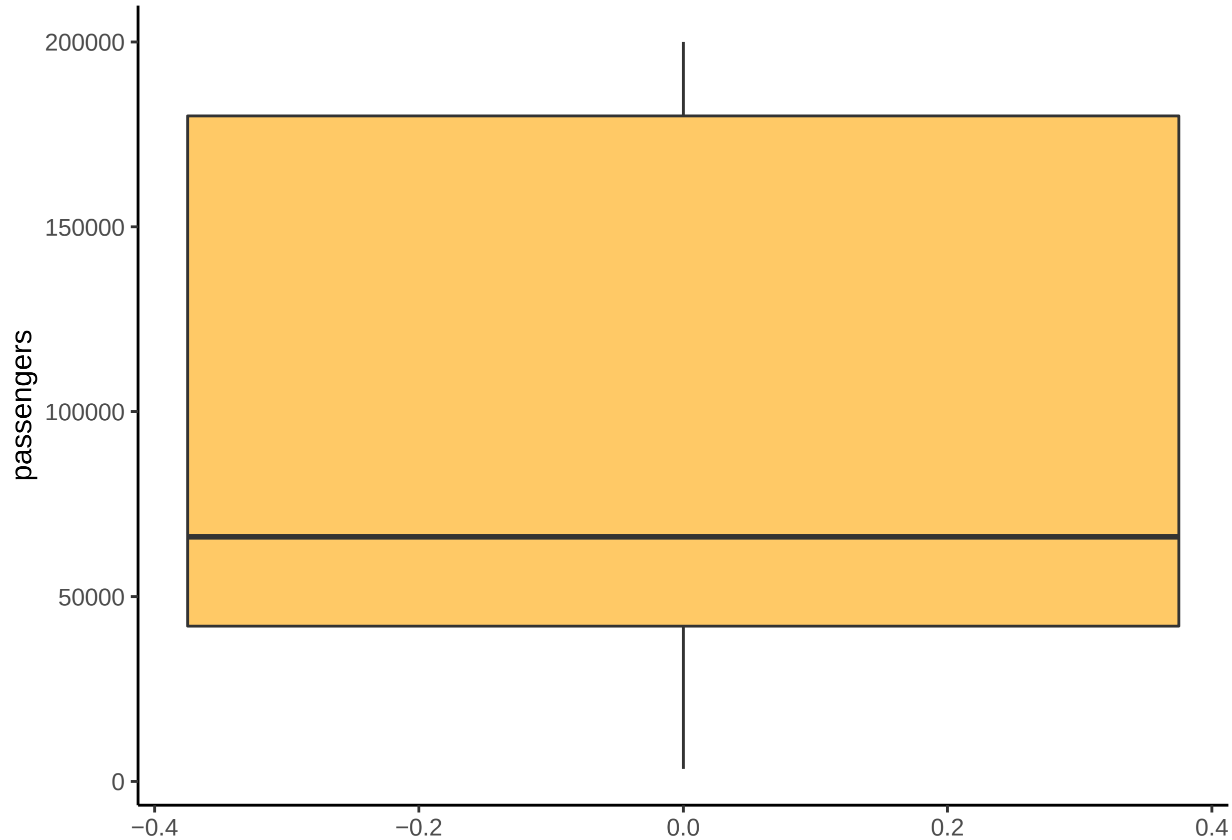
```
##     0%    25%    50%    75%   100%
##   3400  42000  66178 180000 200000
```

```
IQR((afrika$passengers))
```

```
## [1] 138000
```

```
#Visualize quartiles
p2 <- ggplot(afrika,aes(passengers)) +
  geom_boxplot(fill="orange", alpha=0.6)

p2 + theme_classic()+
  coord_flip()
```



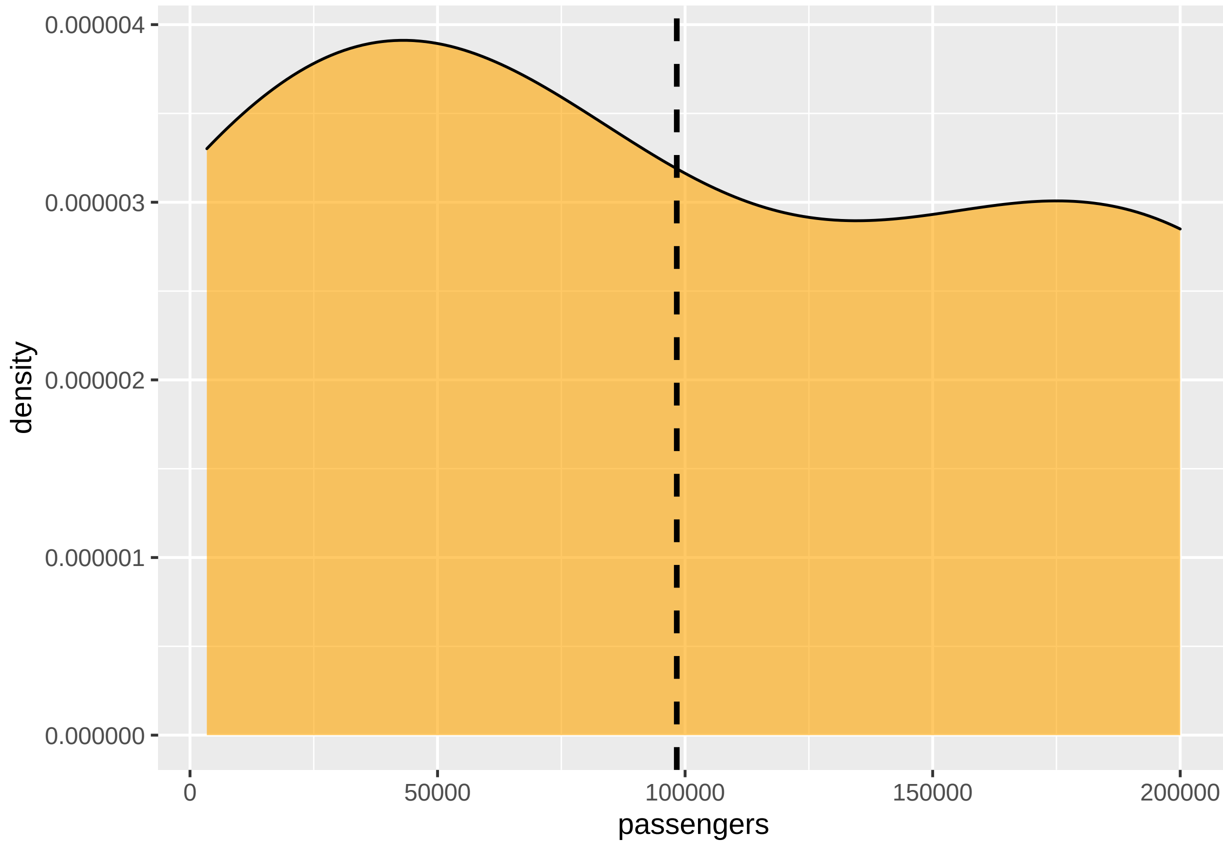
As the boxplot above vividly shows, the distribution of passengers is relatively more uniform with no outliers and a concentration of the data between 42,000 - 180,000 range with an interquartile range of 138,000.

Another interesting point, as the density plot below shows is that, it is worth noting is that, the likeliest number of daily passengers you would get in a given African city based on the dataset is between 25,000 - 75,000 as our density curve below shows. An interesting insight into the data that may be a more reliable/realistic yardstick in future planning of further BRT systems in other countries.

```
#Density plot for Africa
set.seed(250)

p3 <- ggplot(afrika,aes(passengers)) +
  geom_density(fill="orange", alpha=0.6)

p3 + geom_vline(aes(xintercept=mean(afrika$passengers)),
            color="black", linetype="dashed", size=1)+
  scale_y_continuous(labels=comma)
```



Let us now move from Africa further north to Europe. Now, Europe is arguably the continent with the most developed the most developed bus system in the world. No pun intended, but how does the BRT system fare there?

```
#Load datasets for European countries
finland <- read_csv("brtdata-finland.csv")
```

```
## Rows: 1 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
france <- read_csv("brtdata-france.csv")
```

```
## Rows: 19 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
germany <- read_csv("brtdata-germany.csv")
```

```
## Rows: 2 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
netherlands <- read_csv("brtdata-netherlands.csv")
```

```
## Rows: 5 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
spain <- read_csv("brtdata-spain.csv")
```

```
## Rows: 2 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
sweden <- read_csv("brtdata-sweden.csv")
```

```
## Rows: 3 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
switzerland <- read_csv("brtdata-switzerland.csv")
```

```
## Rows: 1 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
```

```
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
turkey <- read_csv("brtdata-turkey.csv")
```

```
## Rows: 1 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
united_kingdom <- read_csv("brtdata-united_kingdom.csv")
```

```
## Rows: 6 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
#Bind datasets into a single dataframe
europa <- rbind(
  finland,france,germany,netherlands,spain,sweden,
  switzerland,turkey,united_kingdom)


#Rename columns for more efficient identification
europa <- europa %>%
  rename(
    city=Cities,
    passengers=`Passengers per Day`,
    corridors=`Number of Corridors`,
    length=`Length (km)`
  )

europa
```

```
## # A tibble: 40 x 4
##    city              passengers corridors length
##    <chr>                  <dbl>     <dbl>  <dbl>
##  1 Helsinki               30000         1     28
##  2 Belfort                32000         1      5
##  3 Caen                   48000         1     16
##  4 Chalon-sur-Saone        5300         1      6
##  5 Douai                   1800         1     34
```

```
##   6 La Rochelle            5600         1      6
##   7 Le Mans                5000         1      7
##   8 Lille               1317000         3     67
##   9 Lorient               45000         1      5
##  10 Lyon                   4700         2     19
## # ... with 30 more rows
```

```
str(europa)
```

```
## spec_tbl_df [40 x 4] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ city      : chr [1:40] "Helsinki" "Belfort" "Caen" "Chalon-sur-Saone" ...
##  $ passengers: num [1:40] 30000 32000 48000 5300 1800 ...
##  $ corridors : num [1:40] 1 1 1 1 1 1 1 3 1 2 ...
##  $ length    : num [1:40] 28 5 16 6 34 6 7 67 5 19 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Cities = col_character(),
##   ..   `Passengers per Day` = col_number(),
##   ..   `Number of Corridors` = col_double(),
##   ..   `Length (km)` = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
#Sum of passengers per day and total corridor distance
tot_eur <- europa %>%
  summarize(
    passengers_per_day=sum(passengers),
    total_distance=sum(length)
  )

tot_eur
```

```
## # A tibble: 1 x 2
##   passengers_per_day total_distance
##                <dbl>          <dbl>
## 1            2906913            838
```

The BRT system in Europe is used by a combined total of 2.9 million passengers on a daily basis. The total distance of all corridors on the continent also adds up to 838km.

```
#Top 3 European cities by daily passenger total
top_cities_europa <- europa %>%
  select(city,passengers,corridors,length) %>%
  arrange(desc(passengers))%>%
  slice(1:3)

top_cities_europa
```

```
## # A tibble: 3 x 4
##   city     passengers corridors length
##   <chr>         <dbl>     <dbl>  <dbl>
## 1 Lille       1317000         3     67
## 2 Istanbul     750000         1     52
## 3 Paris         89500         3     41
```

The French city of Lille boasts of the most daily passengers of the BRT system in Europe with approximately 1.32 million passengers per day. Turkish city Istanbul is the next city with the most BRT passengers with a daily passenger total of 750,000. French capital Paris comes in at third place with 89,500 daily passengers and Stockholm and Caen complete the top 5 European cities according to daily passenger total with 57,000 and 48,000 daily passengers respectively.

Figure 4: Lille

It is in Europe that we finally achieve some normalcy with the top 5 cities per daily passenger total the same as the top 5 cities per total corridor length. Lille and Paris however rank joint first placed based on the number of corridors with 3 each. Turkey's Istanbul's total passengers per day figure looks even more impressive given the fact that, there is only a single corridor running through the entire city.

We have seen that Lille and Istanbul are the cities in Europe with the highest daily passenger totals but wouldn't we get more insight by seeing the average number of daily passengers European BRT system have? Bien sûr!

```
#Calculate mean of the passenger distribution
avg_eu <- europa %>%
  select(city,passengers) %>%
  summarize(
    avg_ppl=mean(passengers),
    avg_km=mean(europa$length)
  )
```

```r
avg_eu
```

```
## # A tibble: 1 x 2
##   avg_ppl avg_km
##     <dbl>  <dbl>
## 1  72673.   21.0
```

Intéressant! Interesting! Europe has an average daily passenger total of 73,000 people, which is relatively very low. They post less figures for passengers per day and total corridor distance than Asia and even Africa, despite having more cities!

```r
#Split to quartiles to check real structure of passenger distribution
quantile(europa$passengers)
```

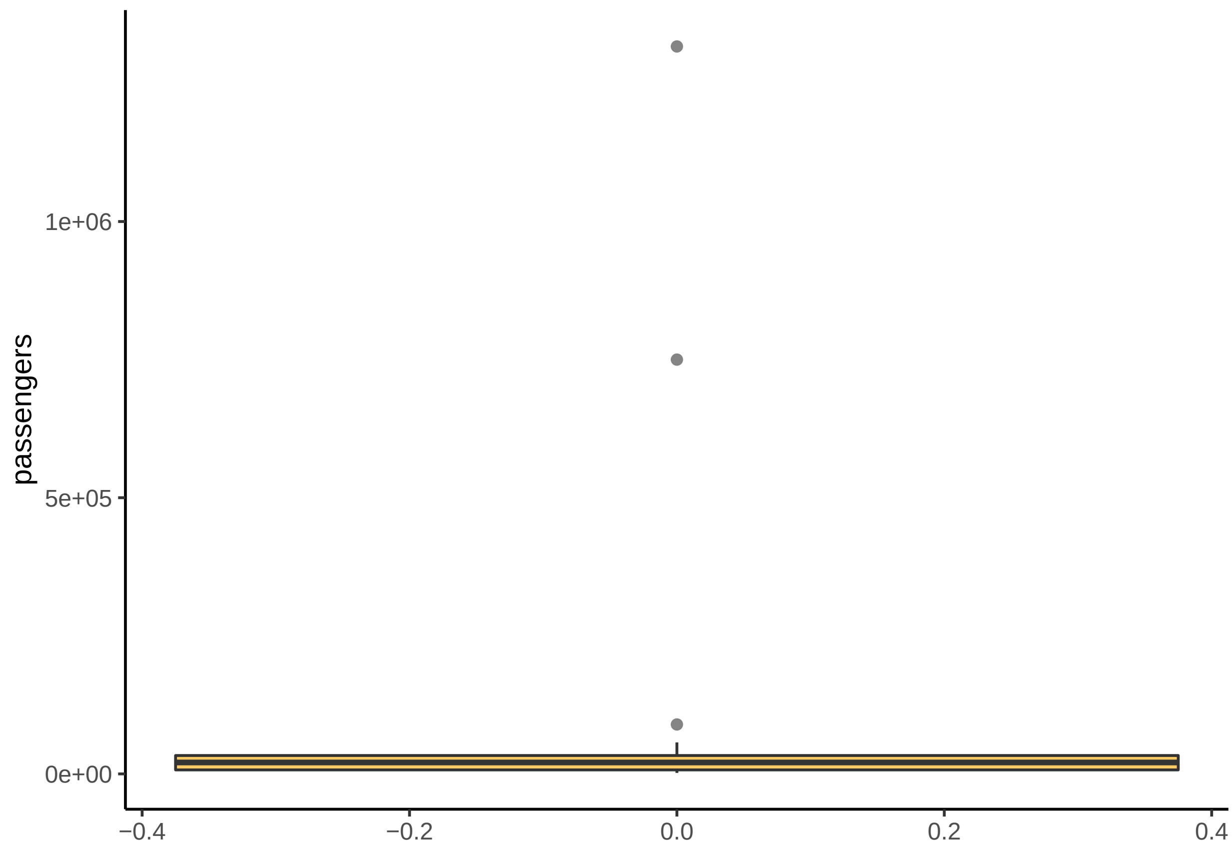
```
##         0%        25%        50%        75%       100%
##    1800.00    7220.75   20666.50   33250.00 1317000.00
```

```r
IQR((europa$passengers))
```

```
## [1] 26029.25
```

```r
#Visualize quartiles
p4 <- ggplot(europa,aes(passengers)) +
  geom_boxplot(fill="orange", alpha=0.6)

p4 + theme_classic()+
  coord_flip()
```



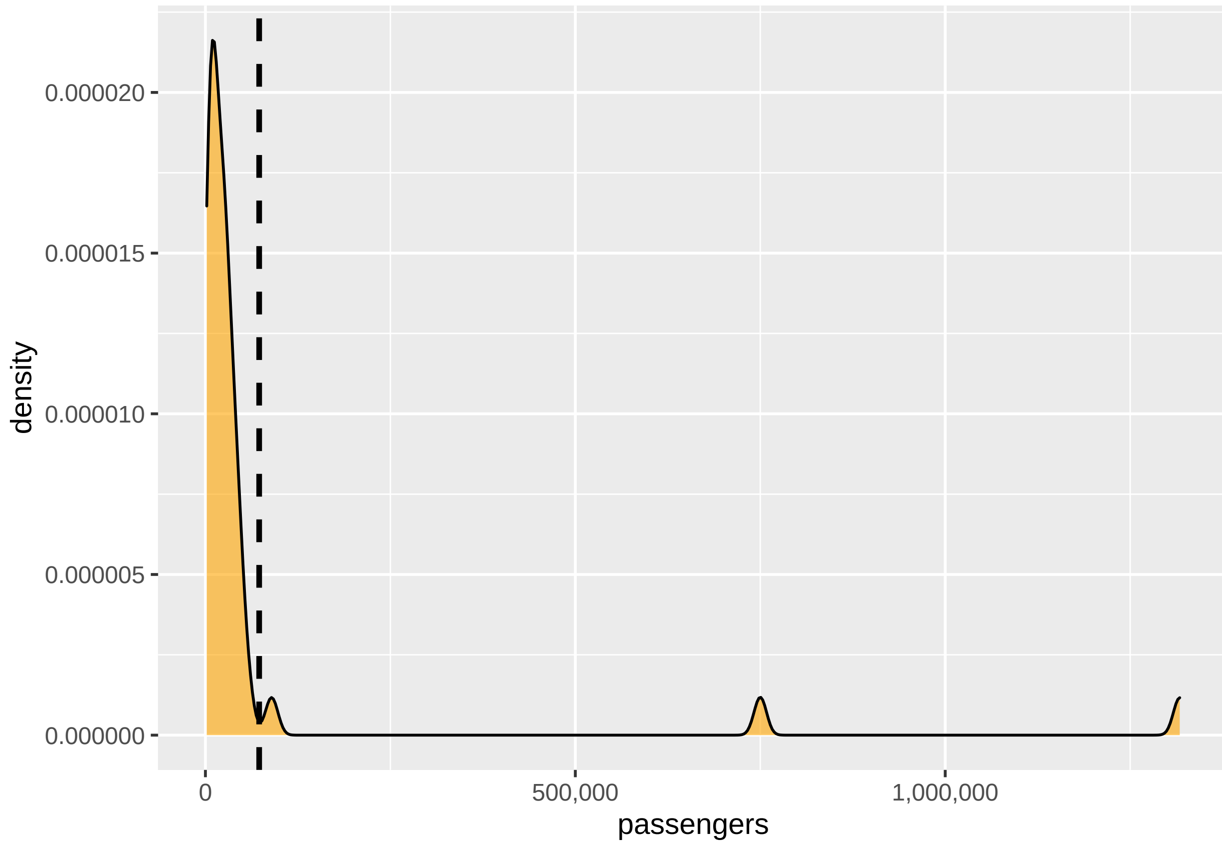
The above boxplot shows us how close together number of daily passengers of BRT systems in Europe are with the exception of the top 3; Lille, Istanbul and Paris, all accounting for the three outliers outside our data. Our interquartile range produces a figure of 26,029, a figure which gives us a very accurate measure of spread than the standard range which will definitely be affected by the outliers.

From our density curve, we see a higher occurrence of cities having daily passenger totals between 7,200 - 33,000. Our density curve below shows us a higher probability of a chosen European city having less than the average number of daily passengers which was around 73,000. This is a more realistic metric to use in assessing the standard for future BRT systems.

```
#Density plot for Europe
set.seed(250)

p5 <- ggplot(europa,aes(passengers)) +
  geom_density(fill="orange", alpha=0.6)

p5 + geom_vline(aes(xintercept=mean(europa$passengers)),
            color="black", linetype="dashed", size=1)+
  scale_x_continuous(labels=comma)+
  scale_y_continuous(labels=comma)
```



Our next point of focus will be across the Atlantic Ocean from Europe straight to North America! Now there are two countries under scrutiny here; Canada and the United States of America. Will their figures be better than the previous 3 continents? Let's get right into it!

```
#Load datassets
canada <- read_csv("brtdata-canada.csv")
```

```
## Rows: 6 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
united_states <- read_csv("brtdata-united_states.csv")
```

```
## Rows: 14 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
#Bind datasets into a single dataframe
nord_amerika <- rbind(
  canada,united_states
)

#Rename columns for more efficient identification
nord_amerika <- nord_amerika %>%
  rename(
    city=Cities,
    passengers=`Passengers per Day`,
    corridors=`Number of Corridors`,
    length=`Length (km)`
  )

nord_amerika
```

```
## # A tibble: 20 x 4
##    city                         passengers corridors length
##    <chr>                             <dbl>     <dbl>  <dbl>
##  1 Calgary                            7640         4     56
##  2 Mississauga                       15100         1     18
##  3 Ottawa                           220000         5     59
##  4 Quebec                            85310         4    120
##  5 Winnipeg                         166000         1      4
##  6 York Regional Municipality         9357         1     38
##  7 Albuquerque                        8100         2     27
##  8 Alexandria - Arlington             1412         1     11
##  9 Cleveland                         15000         1     11
## 10 Eugene                            10000         2     19
## 11 Fort Collins                       3000         1      8
## 12 Hartford County                    9674         1     15
## 13 Las Vegas                         34189         3     63
## 14 Los Angeles                       26179         2     74
## 15 Miami                             23000         1     34
## 16 New York                         305911        12    102
## 17 Orlando                            1588         3      5
## 18 Pittsburgh                        41861         3     30
## 19 San Bernardino                     2475         1     26
## 20 San Francisco                     20000         1      0
```

```
str(nord_amerika)
```

```
## spec_tbl_df [20 x 4] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ city      : chr [1:20] "Calgary" "Mississauga" "Ottawa" "Quebec" ...
##  $ passengers: num [1:20] 7640 15100 220000 85310 166000 ...
##  $ corridors : num [1:20] 4 1 5 4 1 1 2 1 1 2 ...
##  $ length    : num [1:20] 56 18 59 120 4 38 27 11 11 19 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Cities = col_character(),
##   ..   `Passengers per Day` = col_number(),
##   ..   `Number of Corridors` = col_double(),
##   ..   `Length (km)` = col_double()
```

```
##   .. )
##  - attr(*, "problems")=<externalptr>
```

So our North America dataframe contains 20 cities, 6 Canadian cities and 14 American cities.

```
#Sum of passengers per day and total corridor distance
tot_na <- nord_amerika %>%
  summarize(
    passengers_per_day=sum(passengers),
    total_distance=sum(length)
  )

tot_na
```

```
## # A tibble: 1 x 2
##   passengers_per_day total_distance
##                <dbl>          <dbl>
## 1            1005796            720
```

A total of a little over 1 million passengers use the BRT system in North America, less than Europe and Asia but more than Africa. The total distance of all corridors in the region is 720km.

```
#Top 3 cities by daily passenger total
top_cities_n_amerika <- nord_amerika %>%
  select(city,passengers,corridors,length) %>%
  arrange(desc(passengers))%>%
  slice(1:3)

top_cities_n_amerika
```

```
## # A tibble: 3 x 4
##   city     passengers corridors length
##   <chr>         <dbl>     <dbl>  <dbl>
## 1 New York     305911        12    102
## 2 Ottawa       220000         5     59
## 3 Winnipeg     166000         1      4
```

America's New York, the city with the most corridors and second most corridor length total, comes in first position based on the total number of daily passengers of BRT systems. 305,911 New Yorkers use BRT on a daily basis. Canada's Ottawa comes in next 220,000 daily passengers over 5 corridors that spans 59km in all.

Winnipeg however, has one very interesting statistic. Yes, it is the third city in North America with the highest number of daily passengers but it only has a single corridor that is only 4km long! It boasts of the highest number of daily passengers over the shortest distance in the world!

```
#Calculate mean of the passenger distribution
avg_na <- nord_amerika %>%
  select(city,passengers) %>%
  summarize(
    avg_ppl=mean(passengers),
    avg_km=mean(nord_amerika$length)
  )

avg_na
```

Figure 5: New York

```
## # A tibble: 1 x 2
##   avg_ppl avg_km
##     <dbl>  <dbl>
## 1  50290.     36
```

Averagely, the daily number of BRT system passengers in North America are only about 50,000 per day. The lowest so far! The average length of its corridors is 36km.

```
#Split to quartiles to check real structure of passenger distribution
quantile(nord_amerika$passengers)
```

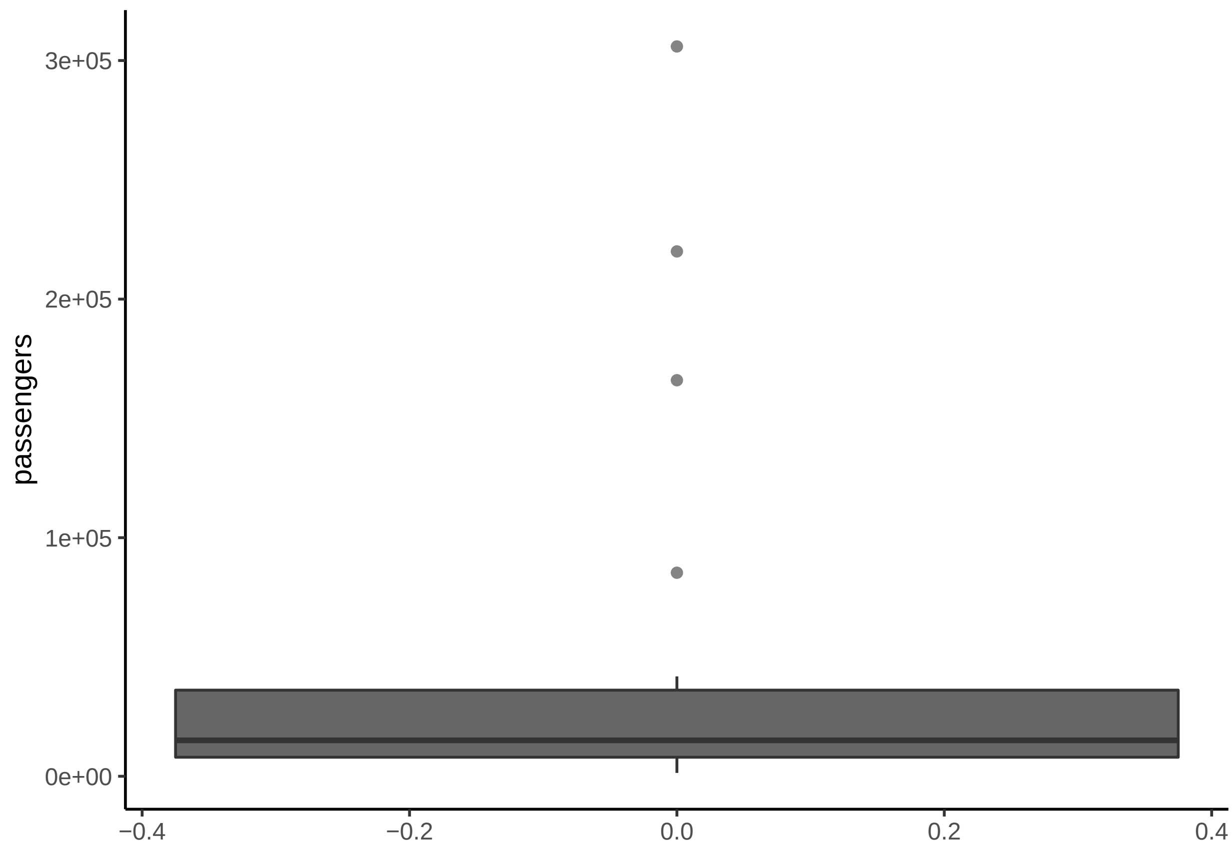
```
##     0%    25%    50%    75%   100%
##   1412   7985  15050  36107 305911
```

```
IQR(nord_amerika$passengers)
```

```
## [1] 28122
```

```
#Visualize quartiles
p6 <- ggplot(nord_amerika,aes(passengers)) +
  geom_boxplot(fill="black", alpha=0.6)

p6 + theme_classic()+
  coord_flip()
```



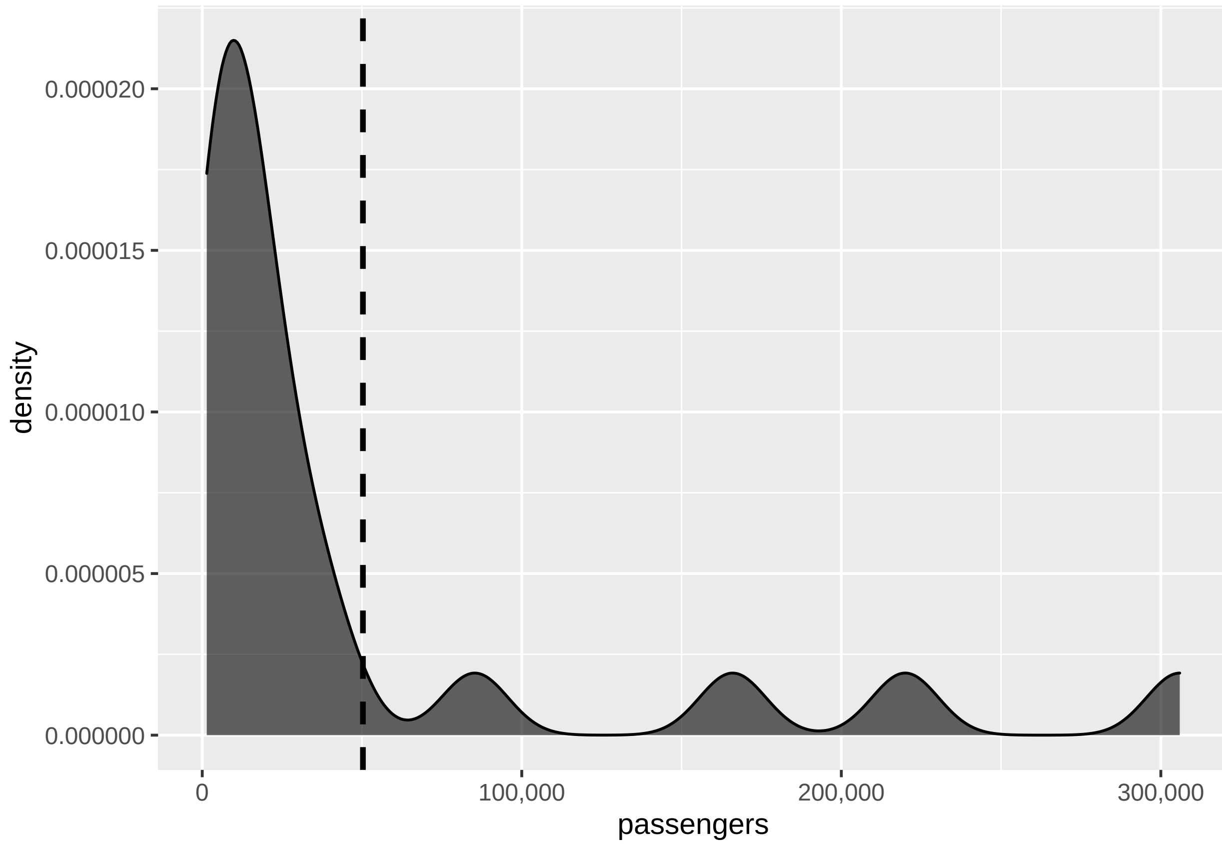
The above boxplot shows us the range of our data with much of the data points/cities with daily passenger numbers below 50,000. The outliers are New York, Ottawa, Winnepeg and Quebec. The majority of the data lies between 8,000 to 36,000, showing an interquartile range of around 28,000.

Our density plot confirms the above notion as a larger amount of cities have a daily passenger total of less than 28,000

```r
#Density plot for North America
set.seed(250)

p7 <- ggplot(nord_amerika,aes(passengers)) +
  geom_density(fill="black", alpha=0.6)

p7 + geom_vline(aes(xintercept=mean(nord_amerika$passengers)),
            color="black", linetype="dashed", size=1)+
  scale_x_continuous(labels=comma)+
  scale_y_continuous(labels=comma)
```




```r
#Load datassets
argentina <- read_csv("brtdata-argentina.csv")
```

```
## Rows: 5 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
brazil <- read_csv("brtdata-brazil.csv")
```

```
## Rows: 24 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
chile <- read_csv("brtdata-chile.csv")
```

```
## Rows: 2 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
colombia <- read_csv("brtdata-colombia.csv")
```

```
## Rows: 7 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
ecuador <- read_csv("brtdata-ecuador.csv")
```

```
## Rows: 2 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
el_salvador <- read_csv("brtdata-el_salvador.csv")
```

```
## Rows: 1 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
guatemala <- read_csv("brtdata-guatemala.csv")
```

```
## Rows: 1 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
```

```
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
mexico <- read_csv("brtdata-mexico.csv")
```

```
## Rows: 12 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
peru <- read_csv("brtdata-peru.csv")
```

```
## Rows: 1 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
uruguay <- read_csv("brtdata-uruguay.csv")
```

```
## Rows: 1 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
venezuela <- read_csv("brtdata-venezuela.csv")
```

```
## Rows: 3 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
#Bind datasets into a single dataframe
latein_amerika <- rbind(
  argentina,brazil,chile,colombia,ecuador,el_salvador,guatemala,mexico,
  peru,uruguay,venezuela
)

#Rename columns for more efficient identification
latein_amerika <- latein_amerika %>%
  rename(
    city=Cities,
    passengers=`Passengers per Day`,
    corridors=`Number of Corridors`,
    length=`Length (km)`
  )

latein_amerika
```

```
## # A tibble: 59 x 4
##    city                               passengers corridors length
##    <chr>                                   <dbl>     <dbl>  <dbl>
##  1 Buenos Aires                           522000         9     62
##  2 Buenos Aires - Metropolitan Area       904000         6     43
##  3 Cordoba                                 58000         1      5
##  4 Neuquen                                 35000         1      6
##  5 Santa Fe                                44000         1      6
##  6 Belem                                  100000         1      6
##  7 Belo Horizonte                         248922         7     39
##  8 Brasilia                                51000         2     49
##  9 Campinas                               295000         3     64
## 10 Criciuma                                10470         1      8
## # ... with 49 more rows
```

```r
str(latein_amerika)
```

```
## spec_tbl_df [59 x 4] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ city      : chr [1:59] "Buenos Aires" "Buenos Aires - Metropolitan Area" "Cordoba" "Neuquen" ...
##  $ passengers: num [1:59] 522000 904000 58000 35000 44000 ...
##  $ corridors : num [1:59] 9 6 1 1 1 1 7 2 3 1 ...
##  $ length    : num [1:59] 62 43 5 6 6 6 39 49 64 8 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Cities = col_character(),
##   ..   `Passengers per Day` = col_number(),
##   ..   `Number of Corridors` = col_double(),
##   ..   `Length (km)` = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

There are 59 cities in our Latin America dataframe. 5 Argentinian, 24 Brazilian, 2 Chilean, 7 Colombian, 2 Ecuadorian, 1 El Salvadorian, 1 Guatemalan, 12 Mexican, 1 Peruvian, 1 Uruguayan and 3 Venezuelan cities in the dataframe.

```
#Sum of passengers per day and total corridor distance for Latin America
tot_la <- latein_amerika %>%
  summarize(
    passengers_per_day=sum(passengers),
    total_distance=sum(length)
  )

tot_la
```

```
## # A tibble: 1 x 2
##   passengers_per_day total_distance
##                <dbl>          <dbl>
## 1           17500993           1960
```

A total of a whooping 17.5 million Latin Americans use the BRT system every single day. The total corridor length is also huge with the entire corridor network on the continent adding up to 1960km. The highest performer in terms of both indicators so far, surpassing even Asia and interestingly Africa, Europe and North America combined in terms of combined corridor distances and all 4 continents in terms of daily passengers.

```
#Top 3 Latin American cities by passengers per day
top_cities_l_amerika <- latein_amerika %>%
  select(city,passengers,corridors,length) %>%
  arrange(desc(passengers))%>%
  slice(1:3)

top_cities_l_amerika
```

```
## # A tibble: 3 x 4
##   city           passengers corridors length
##   <chr>               <dbl>     <dbl>  <dbl>
## 1 Rio de Janeiro    3535466        17    168
## 2 Bogota            2192009        11    113
## 3 Mexico City       1240000         7    140
```

In terms of top 3 Latin American countries by daily passengers, Brazil's Rio de Janeiro comes top of the pops with 3.5 million daily. It is the city with the highest number of daily passengers so far. Rio has 17 corridors, spanning a total distance of 168km. Next up is Colombia's Bogotá which boasts of approximately 2.2 million daily passengers wit 11 corridors spanning a combined distance of 113km. Mexico City comes in third place with 1.2 million daily passengers and 7 corridors spanning 140km.

It is interesting to also note that so far, only Latin America has posted over a million figures for its top 3 cities based on total daily passengers.

```
#Calculate mean of the passenger distribution
avg_la <- latein_amerika %>%
  select(city,passengers) %>%
  summarize(
    avg_ppl=mean(passengers),
    avg_km=mean(latein_amerika$length)
  )

avg_la
```

Figure 6: Rio de Janeiro

```
## # A tibble: 1 x 2
##   avg_ppl avg_km
##     <dbl>  <dbl>
## 1  296627   33.2
```

An average number of around 297,000 of Latin Americans use the BRT system on a daily basis. This figure looks even more impressive when you consider the fact that, so far, they are the third placed continent in terms of average distance behind Asia and North America.

```
#Split to quartiles to check real structure of passenger distribution
quantile(latein_amerika$passengers)
```

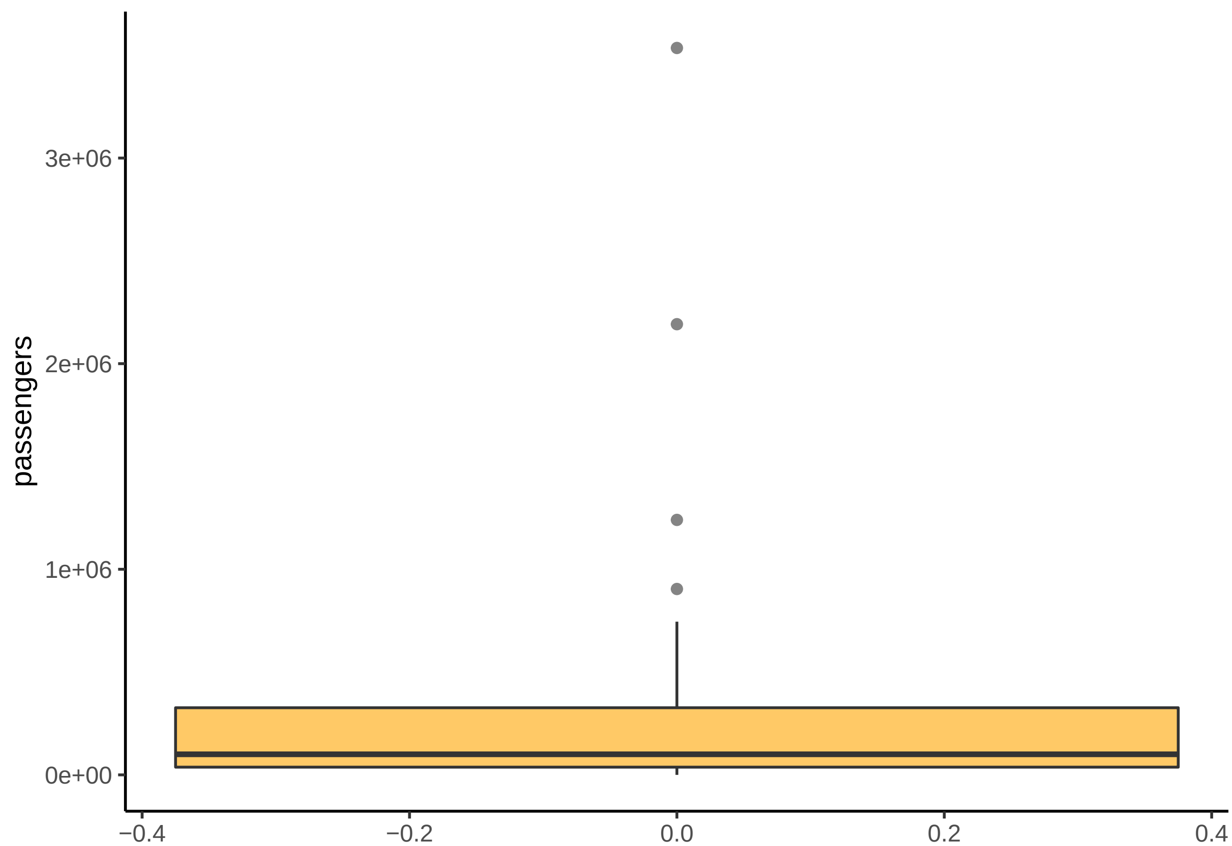
```
##      0%     25%     50%     75%    100%
##       0   37500  100000  326500 3535466
```

```
IQR(latein_amerika$passengers)
```

```
## [1] 289000
```

```
#Visualize quartiles
p8 <- ggplot(latein_amerika,aes(passengers)) +
  geom_boxplot(fill="orange", alpha=0.6)

p8 + theme_classic()+
  coord_flip()
```



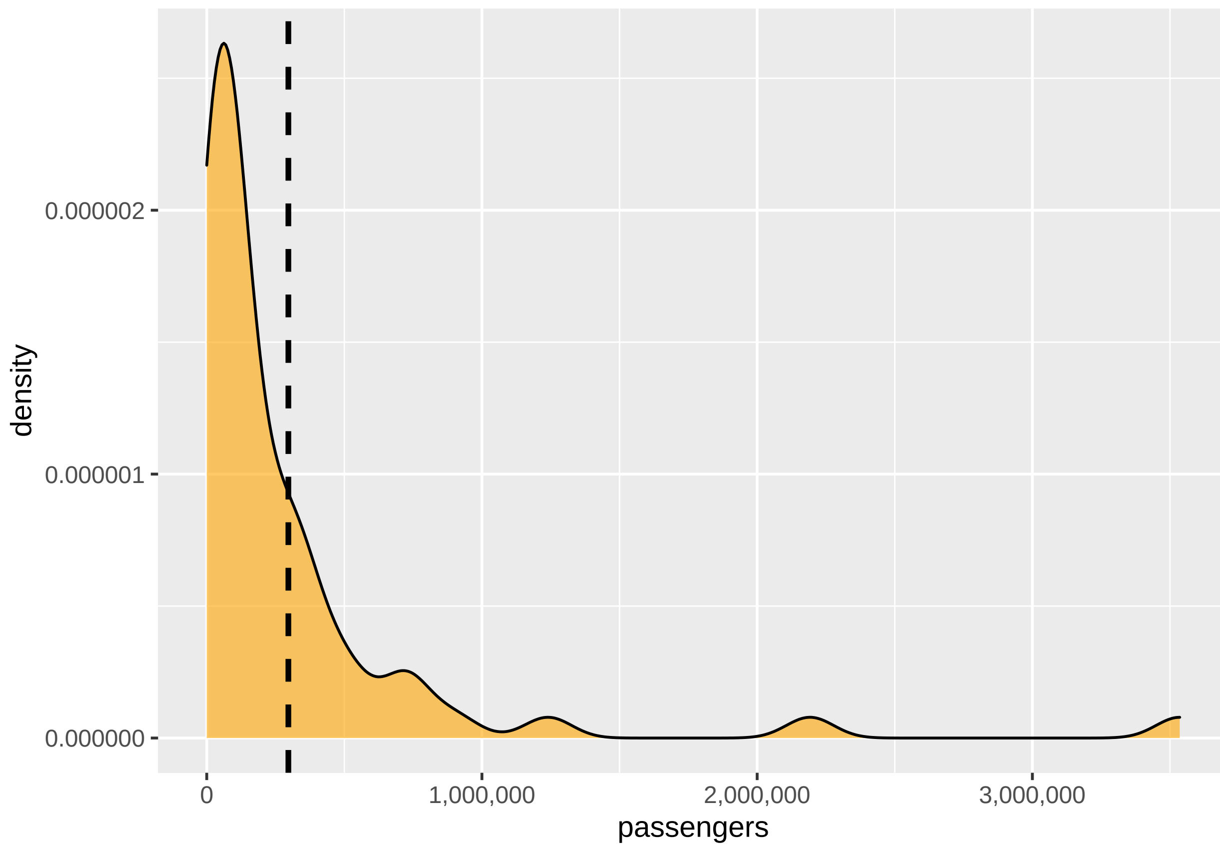
Latin American cities generally have a total of daily passengers between 37,500 - 326,500 as the boxplot above shows. with an interquartile range of 289,000.

The density plot below shows a large number of Latin American countries, more than half in fact below 250,000. The highest interquartile range out of all the continents!

```
#Density plot for Latin Amerika
set.seed(250)

p9 <- ggplot(latein_amerika,aes(passengers)) +
  geom_density(fill="orange", alpha=0.6)

p9 + geom_vline(aes(xintercept=mean(latein_amerika$passengers)),
            color="black", linetype="dashed", size=1)+
  scale_x_continuous(labels=comma)+
  scale_y_continuous(labels=comma)
```



Finally, we analyse Oceania! A continent in the ocean, no?

```
#Load datassets
australia <- read_csv("brtdata-australia.csv")
```

```
## Rows: 3 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
```

```
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
new_zealand <- read_csv("brtdata-new_zealand.csv")
```

```
## Rows: 1 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Cities
## dbl (2): Number of Corridors, Length (km)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
#Bind datasets into a single dataframe
oceania <- rbind(
  australia,new_zealand
)

#Rename columns for more efficient identification
oceania <- oceania %>%
  rename(
    city=Cities,
    passengers=`Passengers per Day`,
    corridors=`Number of Corridors`,
    length=`Length (km)`
  )

oceania
```

```
## # A tibble: 4 x 4
##   city                          passengers corridors length
##   <chr>                              <dbl>     <dbl>  <dbl>
## 1 Adelaide                           32000         1     12
## 2 Brisbane                          356800         3     28
## 3 Sydney - Metropolitan Area         24500         3     49
## 4 Auckland                           22900         1      6
```

```
str(oceania)
```

```
## spec_tbl_df [4 x 4] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ city      : chr [1:4] "Adelaide" "Brisbane" "Sydney - Metropolitan Area" "Auckland"
##  $ passengers: num [1:4] 32000 356800 24500 22900
##  $ corridors : num [1:4] 1 3 3 1
##  $ length    : num [1:4] 12 28 49 6
##  - attr(*, "spec")=
##   .. cols(
##   ..   Cities = col_character(),
##   ..   `Passengers per Day` = col_number(),
##   ..   `Number of Corridors` = col_double(),
```

```
##   ..   `Length (km)` = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

There are 4 cities in our Oceania dataframe. 3 Australian and 1 city from New Zealand.

```r
#Total passengers per day and total distance for Oceania
tot_oce <- oceania %>%
  summarize(
    passengers_per_day=sum(passengers),
    total_distance=sum(length)
  )

tot_oce
```

```
## # A tibble: 1 x 2
##   passengers_per_day total_distance
##                <dbl>          <dbl>
## 1             436200             95
```

Oceania has a total number of 436,200 daily BRT system passengers and a total corridor distance of 95km.

```r
#Top 3 cities by daily total passenger total
top_cities_oceania <- oceania %>%
  select(city,passengers,corridors,length) %>%
  arrange(desc(passengers))%>%
  slice(1:3)

top_cities_oceania
```

```
## # A tibble: 3 x 4
##   city                         passengers corridors length
##   <chr>                             <dbl>     <dbl>  <dbl>
## 1 Brisbane                         356800         3     28
## 2 Adelaide                          32000         1     12
## 3 Sydney - Metropolitan Area        24500         3     49
```

Oceania's top 3 countries according to daily passengers are; Brisbane, Adelaide & Sydney with a total daily passenger number of 356,800, 32,000 and 24,500 respectively.

```r
#Calculate mean of the passenger distribution
avg_oce <- oceania %>%
  select(city,passengers) %>%
  summarize(
    avg_ppl=mean(passengers),
    avg_km=mean(oceania$length)
  )

avg_oce
```

```
## # A tibble: 1 x 2
##   avg_ppl avg_km
##     <dbl>  <dbl>
## 1  109050   23.8
```

Figure 7: Brisbane

Oceania has an average number of approximately 109,000 daily passengers and on average, a corridor distance of 23.75km

```
#Split to quartiles to check real structure of passenger distribution
quantile(oceania$passengers)
```

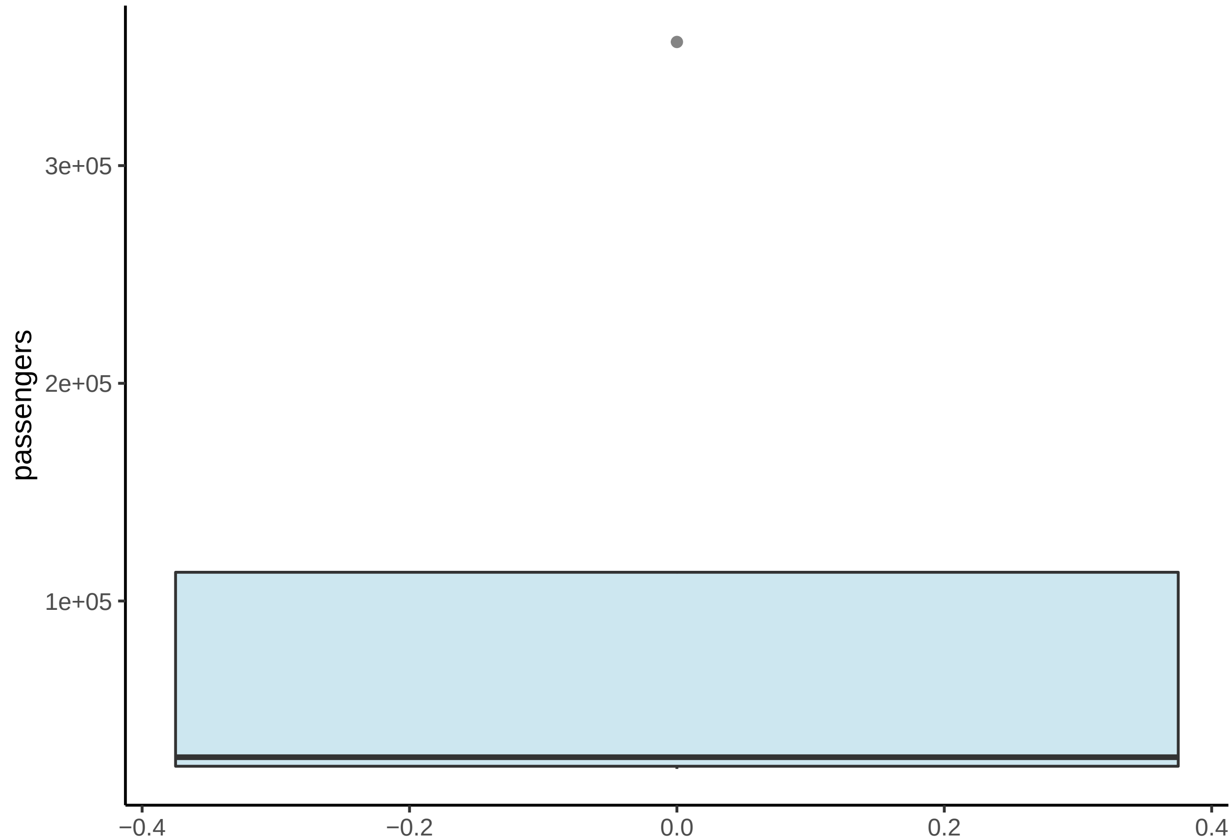
```
##     0%    25%    50%    75%   100%
##  22900  24100  28250 113200 356800
```

```
IQR(oceania$passengers)
```

```
## [1] 89100
```

```
#Visualize quartiles
p10 <- ggplot(oceania,aes(passengers)) +
  geom_boxplot(fill="lightblue", alpha=0.6)

p10 + theme_classic()+
  coord_flip()
```



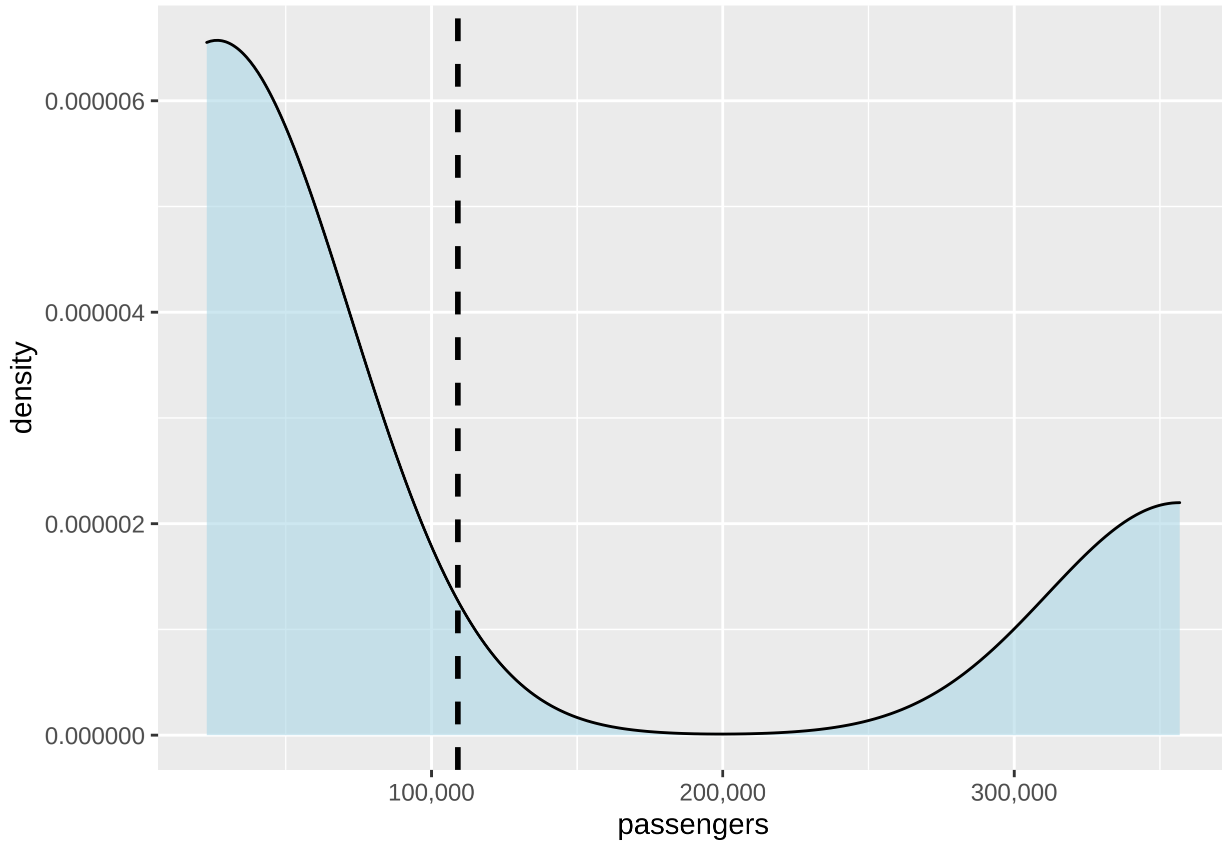
The bulk spread of Oceania's daily passenger total is between 24,100 - 113,200, with an interquartile range of 89,100.

The density plot below shows how the probability of a city's total passenger limit being below the 40,000 mark.

```
#Density plot for Oceania
set.seed(250)

p11 <- ggplot(oceania,aes(passengers)) +
  geom_density(fill="lightblue", alpha=0.6)

p11 + geom_vline(aes(xintercept=mean(oceania$passengers)),
            color="black", linetype="dashed", size=1)+
  scale_x_continuous(labels=comma)+
  scale_y_continuous(labels=comma)
```



We can now test for top continents and cities by; passengers per day, number of corridors and corridor distance. And add a couple of other tests. This analysis will then conclude with a test to check if the number of daily passengers is linked to total number of corridors and corridor length using a regression analysis model.

We would begin with binding all the continent dataframes together.

```
#Bind continents to a single dataframe
kontinenten <- rbind(
  afrika,asien,europa,nord_amerika,latein_amerika,oceania
)

kontinenten
```

```
## # A tibble: 168 x 4
##    city          passengers corridors length
##    <chr>              <dbl>     <dbl>  <dbl>
##  1 Lagos             200000         1     22
##  2 Cape Town          66178         2     31
##  3 Johannesburg       42000         2     44
##  4 Pretoria            3400         2     14
##  5 Dar es Salaam     180000         1     21
##  6 Beijing           305000         4     75
##  7 Changzhou         350000         2     50
##  8 Chongqing          12000         1     12
##  9 Dalian             87000         1     14
## 10 Guangzhou         850000         1     23
```

```
## # ... with 158 more rows
```

We have a total of 168 cities from 6 continents. The cities by continent are; 5 African, 40 Asian, 40 European, 20 North American, 59 Latin American and 4 Oceanic cities in our dataframe. Let us assign them in R.

```
afr_cities <- 5
asi_cities <- 40
eur_cities <- 40
na_cities <- 20
la_cities <- 59
oce_cities <- 4
```

From the above code chunk, the rank by number of cities with a BRT system is: 1. Latin America 2. Asia & Europe 3. North America 4. Africa 5. Oceania

Let us now wrangle our data to find out our top city by daily passenger total for each continent

```
#Assign a continent column
afrika$continent <- "Africa"
asien$continent <- "Asia"
europa$continent <- "Europe"
latein_amerika$continent <- "Latin America"
nord_amerika$continent <- "North America"
oceania$continent <- "Oceania"
```

```
#Africa
top_city_afr <- afrika %>%
  select(city,passengers,continent) %>%
  arrange(desc(passengers))%>%
  slice(1:1)

#Asia
top_city_asi <- asien %>%
  select(city,passengers,continent) %>%
  arrange(desc(passengers))%>%
  slice(1:1)

#Europa
top_city_eur <- europa %>%
  select(city,passengers,continent) %>%
  arrange(desc(passengers))%>%
  slice(1:1)

#Latein Amerika
top_city_la <- latein_amerika %>%
  select(city,passengers,continent) %>%
  arrange(desc(passengers))%>%
  slice(1:1)

#Nord-Amerika
top_city_na <- nord_amerika %>%
  select(city,passengers,continent) %>%
  arrange(desc(passengers))%>%
  slice(1:1)
```

```
#Oceania
top_city_oce <- oceania %>%
  select(city,passengers,continent) %>%
  arrange(desc(passengers))%>%
  slice(1:1)

#Join results
top_city_kontinent <- rbind(
  top_city_afr,top_city_asi,top_city_eur,top_city_la,top_city_na,top_city_oce
)
```

Let us now order and visualize our results.

```
#Rank continents according to their cities with highest daily passenger total
top_city_kontinent %>%
  group_by(passengers,continent) %>%
  arrange(desc(passengers))
```

```
## # A tibble: 6 x 3
## # Groups:   passengers, continent [6]
##   city           passengers continent
##   <chr>               <dbl> <chr>
## 1 Rio de Janeiro    3535466 Latin America
## 2 Tehran            2000000 Asia
## 3 Lille             1317000 Europe
## 4 Brisbane           356800 Oceania
## 5 New York           305911 North America
## 6 Lagos              200000 Africa
```

As we can see from our results above, Latin America has the city with the highest number of daily passengers with Rio de Janeiro at first position. Asia is second with Iran's Tehran coming in second. Europe takes third position with France's Lille coming third and Oceania's Brisbane, North America's New York and Africa's Lagos coming in 4th, 5th and 6th with 356,800, 305911 and 200,000 respectively.

Now to the bigger fish we can fry! We would now find out the top continents by daily passengers.

```
#Rank continents according to total passengers
top_kontinent <- kontinenten%>%
  group_by(kontinenten$continent) %>%
  summarise(passengers = sum(passengers))%>%
  arrange(desc(passengers))
```

```
## Warning: Unknown or uninitialised column: `continent`.
```

```
top_kontinent
```

```
## # A tibble: 1 x 1
##   passengers
##        <dbl>
## 1   31579540
```

Once again, Latin America ranks first in terms of most daily passengers by continent with 17.5 million daily passengers, nearly twice as much as the second ranked continent, Asia and more than all the other continents combined indicating the heavy use of the BRT system in Latin America. Asia is the second continent with the highest total of daily passengers with 9.2 million daily users. Europe and North America rank 3rd and 4th with about 2.9 and 1 million daily users respectively. Africa and Oceania are bottom at 5th and 6th place with 492,000 and 436,000 daily passengers respectively, for Africa, an area that can be massively improved upon.

```
#Rank total corridor distance by continent
top_dist_kontinent <- kontinenten %>%
  group_by(kontinenten$continent) %>%
  summarise(length = sum(length))%>%
  arrange(desc(length))
```

```
## Warning: Unknown or uninitialised column: ‘continent‘.
```

```
top_dist_kontinent
```

```
## # A tibble: 1 x 1
##   length
##    <dbl>
## 1   5347
```

In terms of corridor length, Latin America ranks first, Asia second, Europe third, North America fourth, Africa fifth and Oceania bottom with 1,960, 1,602, 838, 720, 132 & 95 respectively.

```
#Rank corridor total by continent
top_corridors <- kontinenten%>%
  group_by(kontinenten$continent) %>%
  summarise(corridors = sum(corridors))%>%
  arrange(desc(corridors))
```

```
## Warning: Unknown or uninitialised column: ‘continent‘.
```

```
top_corridors
```

```
## # A tibble: 1 x 1
##   corridors
##       <dbl>
## 1       403
```

One can be forgiven for thinking Latin America is done after their impressive numbers but they once again rank first number of BRT corridors with a total of 191. Asia is second with a total of 94 corridors, Europe 52, North America 50, Africa 8 & Oceania 8 in that order.

```
#Cities with over a million daily passenger total
million_daily <- kontinenten%>%
  group_by(kontinenten$city) %>%
  filter(passengers>1000000)%>%
  arrange(desc(passengers))

million_daily
```

```
## # A tibble: 6 x 5
## # Groups:   kontinenten$city [6]
##   city           passengers corridors length ‘kontinenten$city‘
##   <chr>               <dbl>     <dbl>  <dbl> <chr>
## 1 Rio de Janeiro    3535466        17    168 Rio de Janeiro
## 2 Bogota            2192009        11    113 Bogota
## 3 Tehran            2000000         8    130 Tehran
## 4 Lille             1317000         3     67 Lille
## 5 Taipei            1302832         1     60 Taipei
## 6 Mexico City       1240000         7    140 Mexico City
```

There are also only 4 cities in the world with a daily passenger total of over 1 million. Rio de Janeiro is first and by some distance with 3.5 million, followed by Bogota and Tehran with 2.19 and 2 million daily passenger total respectively. Taipei and Mexico City place fifth and sixth respectively with 1.3 and 1.2 million daily passenger totals.

In this respect, we can also rank South America first with 3 countries out of the top 5 cities with over a million daily passengers plus it has the highest ranking city in this respect. Asia comes in second with two cities; Tehran then Taipei and Europe's Lille comes in 3rd.

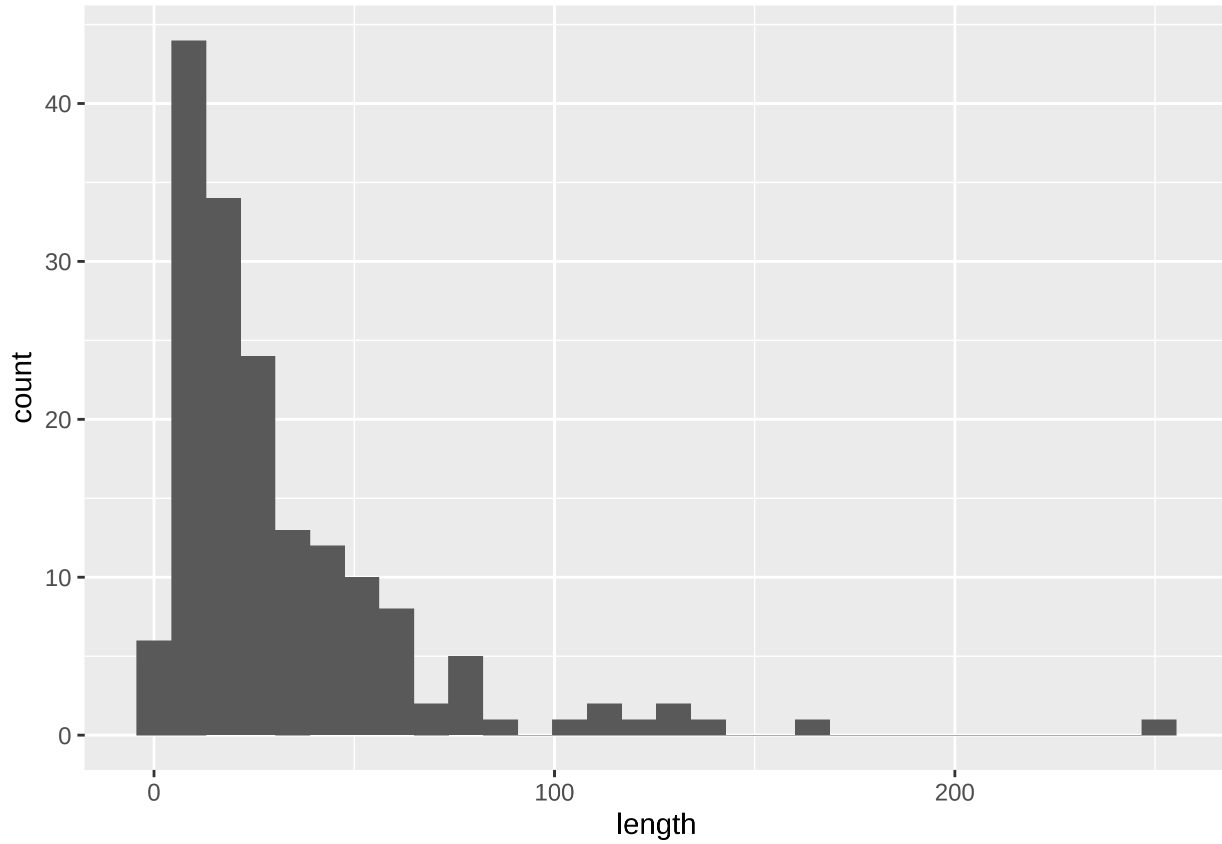
We will now test if there is a correlation between daily passengers and distance and number of corridors respectively.

For our first test between daily passengers and distance, we will state a working and null hypothesis.

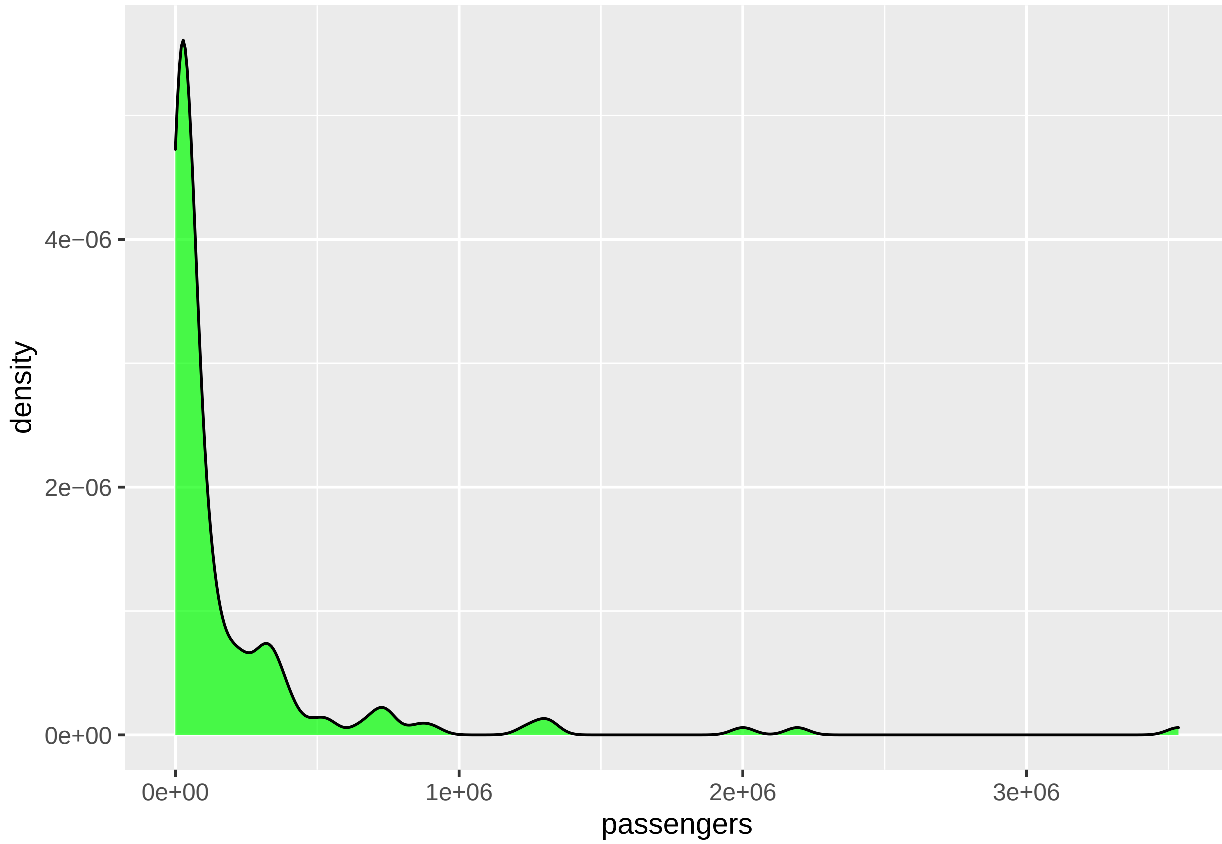
The working(H1) and null(H2) hypotheses are: H1: The higher the distance of corridors, the higher the number of daily passengers H0: There is no link between corridor distance and the number of daily passengers

```r
#Variable 1: length
ggplot(kontinenten,aes(length))+
  geom_histogram()
```

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

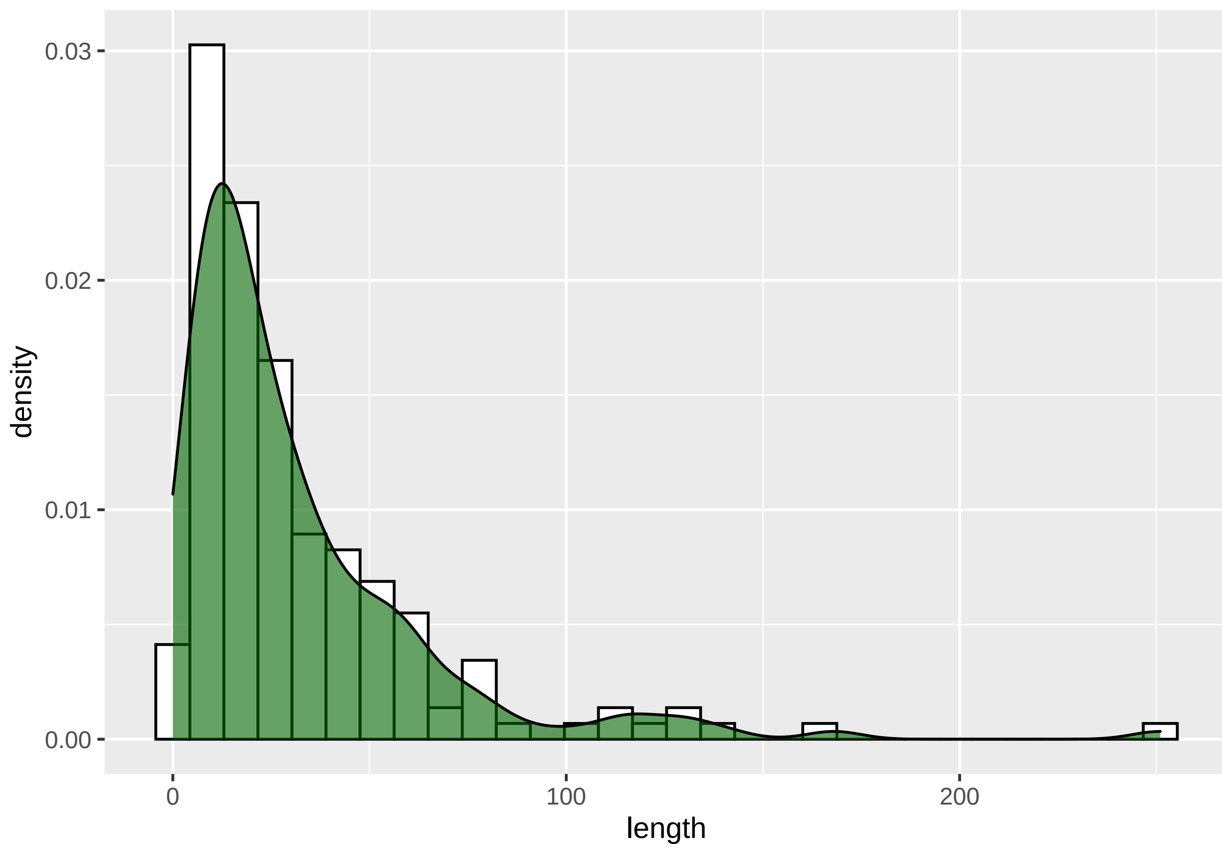



```r
#Variable 2: passengers
ggplot(kontinenten,aes(passengers))+
  geom_density(fill="green",alpha=0.7)
```



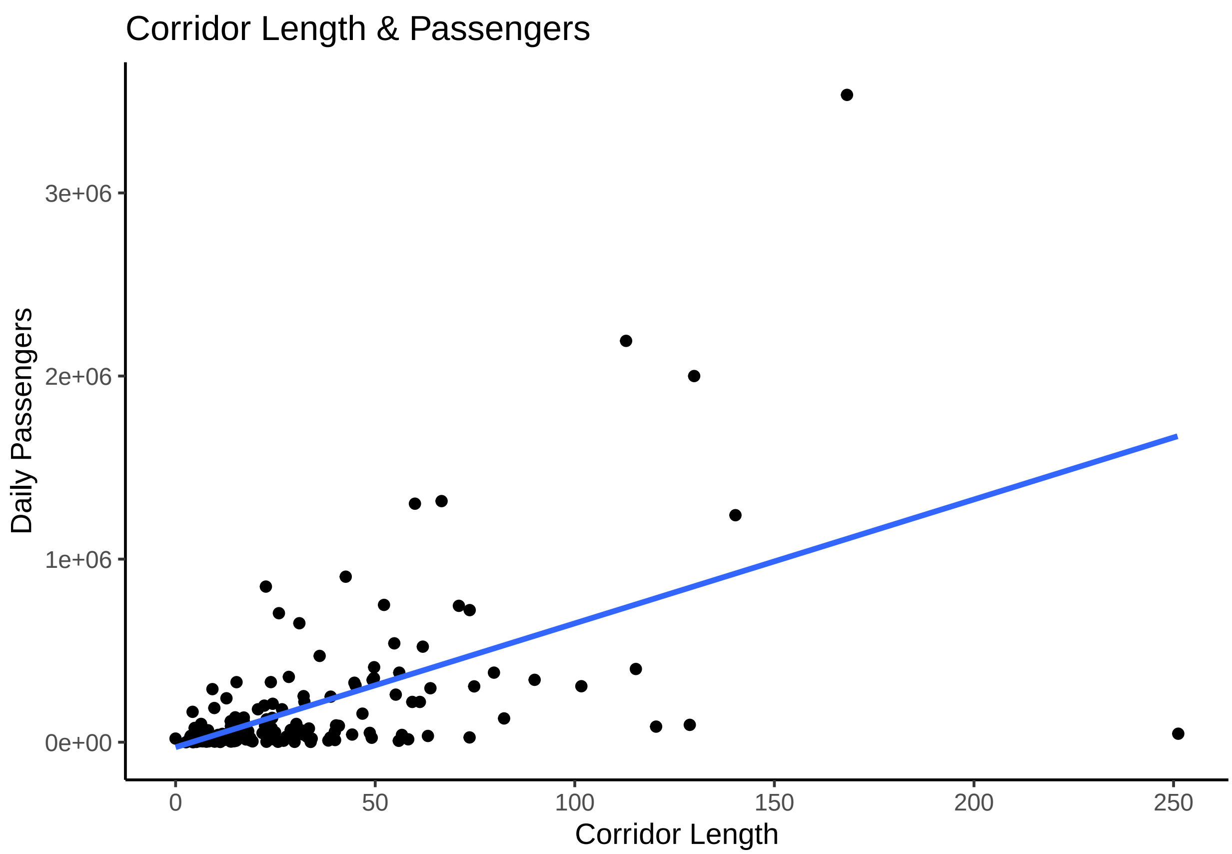

```
#Bivariate relationship visual
ggplot(kontinenten, aes(length)) +
geom_histogram(aes(y=..density..),
colour="black", fill="white") +
geom_density(alpha=0.6, fill="darkgreen")
```

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




```
#Scatterplot to visualize bivariate correlation
ggplot(kontinenten, aes(length,passengers)) +
   geom_point(position='jitter') +
   geom_smooth(method="lm",se=FALSE) +
   labs(title="Corridor Length & Passengers",
      x="Corridor Length",
      y="Daily Passengers") +
   theme_classic()
```

```
## `geom_smooth()` using formula 'y ~ x'
```



The graph above shows the relationship between Corridor Length

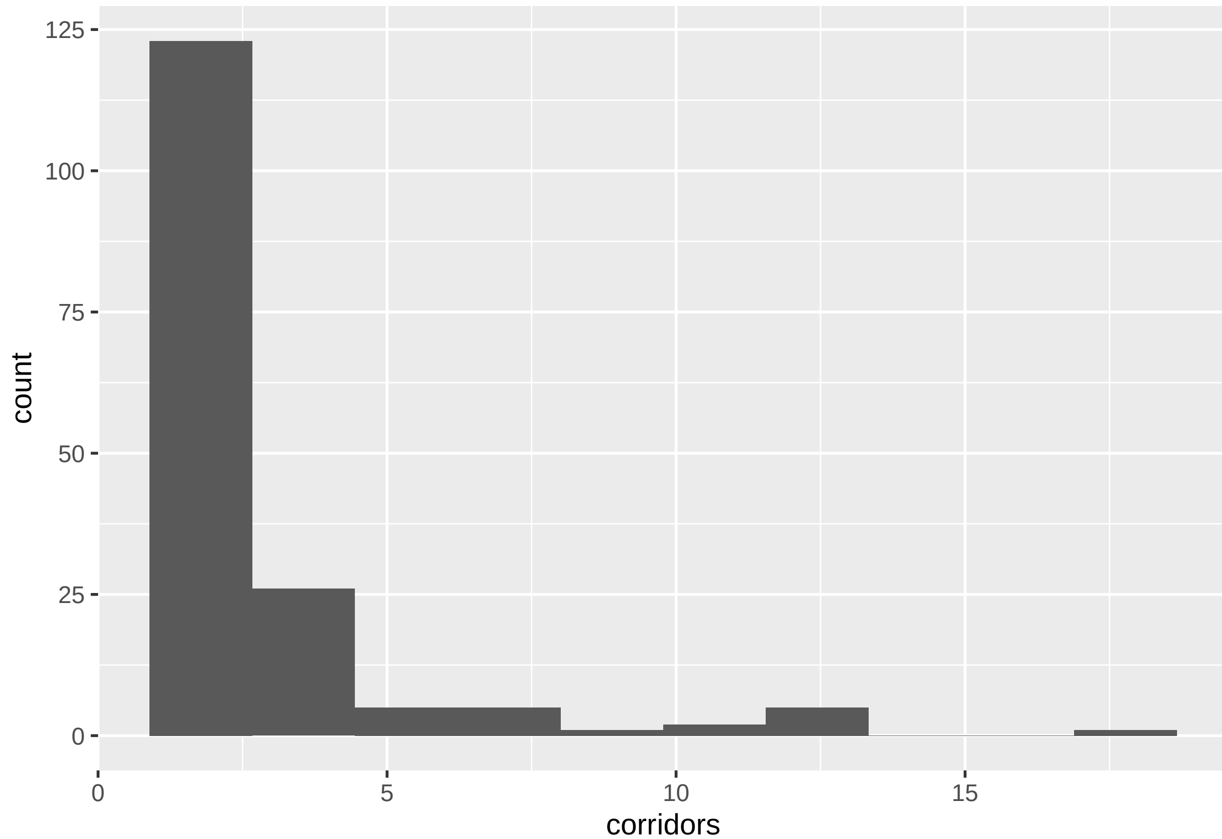
```
#Statistical Correlation test
cor.test(kontinenten$passengers,kontinenten$length)
```

```
##
##  Pearson's product-moment correlation
##
## data:  kontinenten$passengers and kontinenten$length
## t = 8.7016, df = 166, p-value = 3.105e-15
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4460792 0.6555448
## sample estimates:
##       cor
## 0.5596873
```

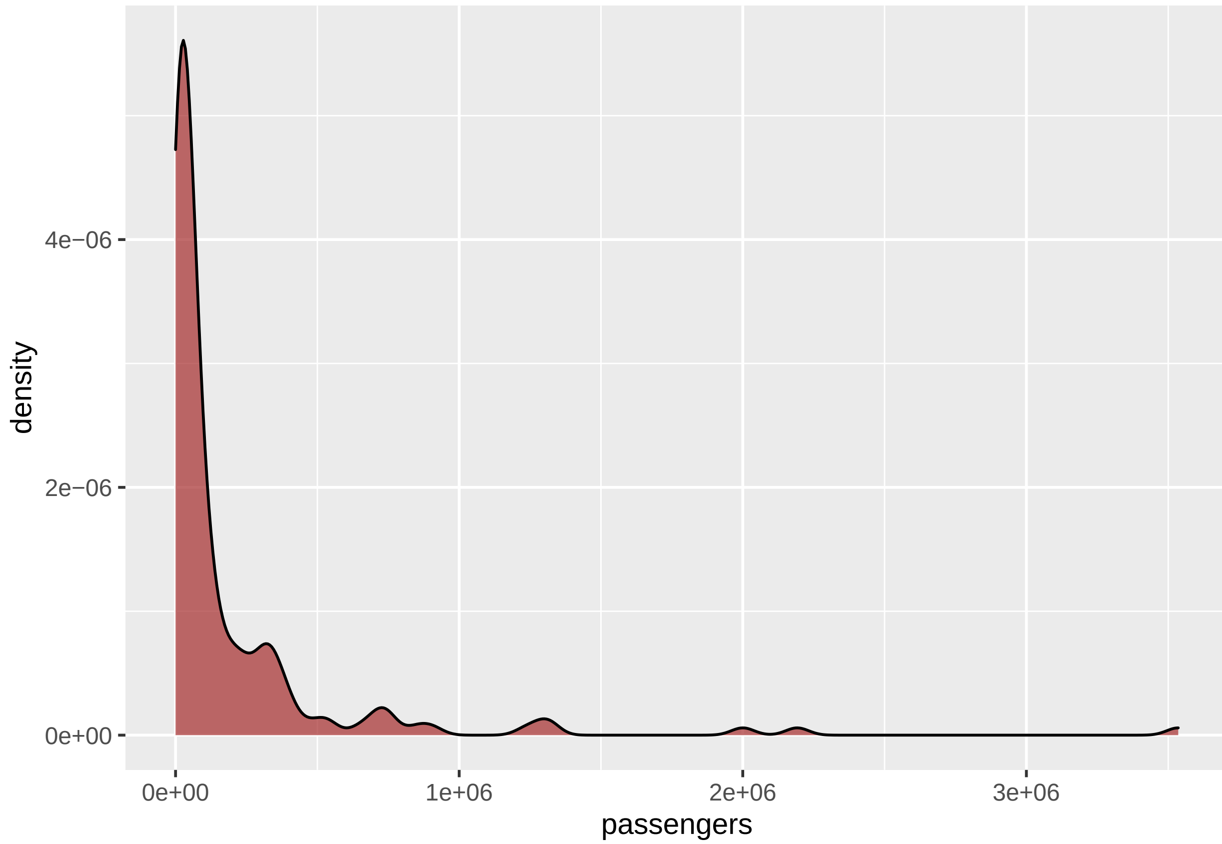
Our Pearson R coefficient is 0.56 with a p-value of 0.0000000000000031. Since our p-value is lower than the 0.05 statistical significance cut-off point, we can conclude that, there is a strong relationship between corridor distance and daily passengers so we can therefore reject the null hypothesis which says "There is no link between corridor distance and the number of daily passengers."

The working(H1) and null(H2) hypotheses are: H1: The higher the number of corridors, the higher the number of daily passengers H0: There is no link between number of corridors and the number of daily passengers

```r
#Variable 1: corridors
ggplot(kontinenten,aes(corridors))+
  geom_histogram(bins=10)
```

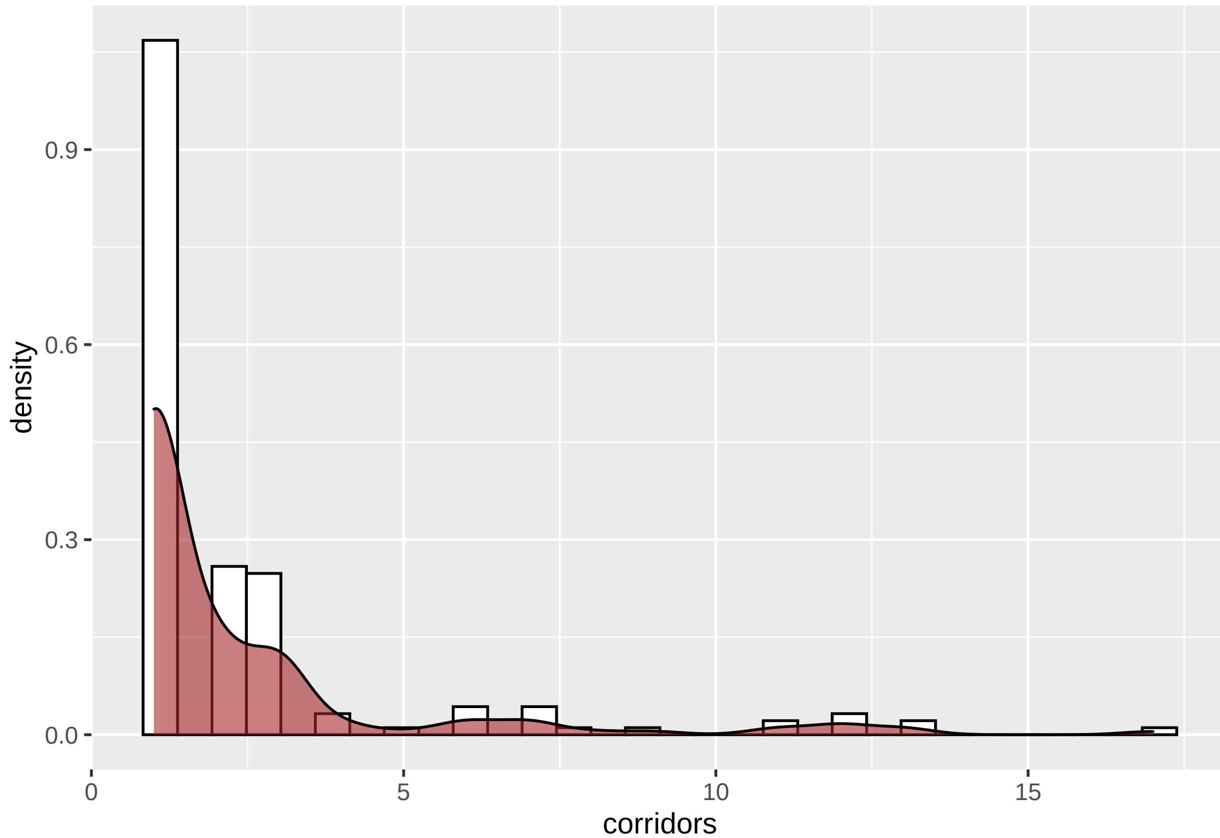



```r
#Variable 2: passengers
ggplot(kontinenten,aes(passengers))+
  geom_density(fill="brown",alpha=0.7)
```



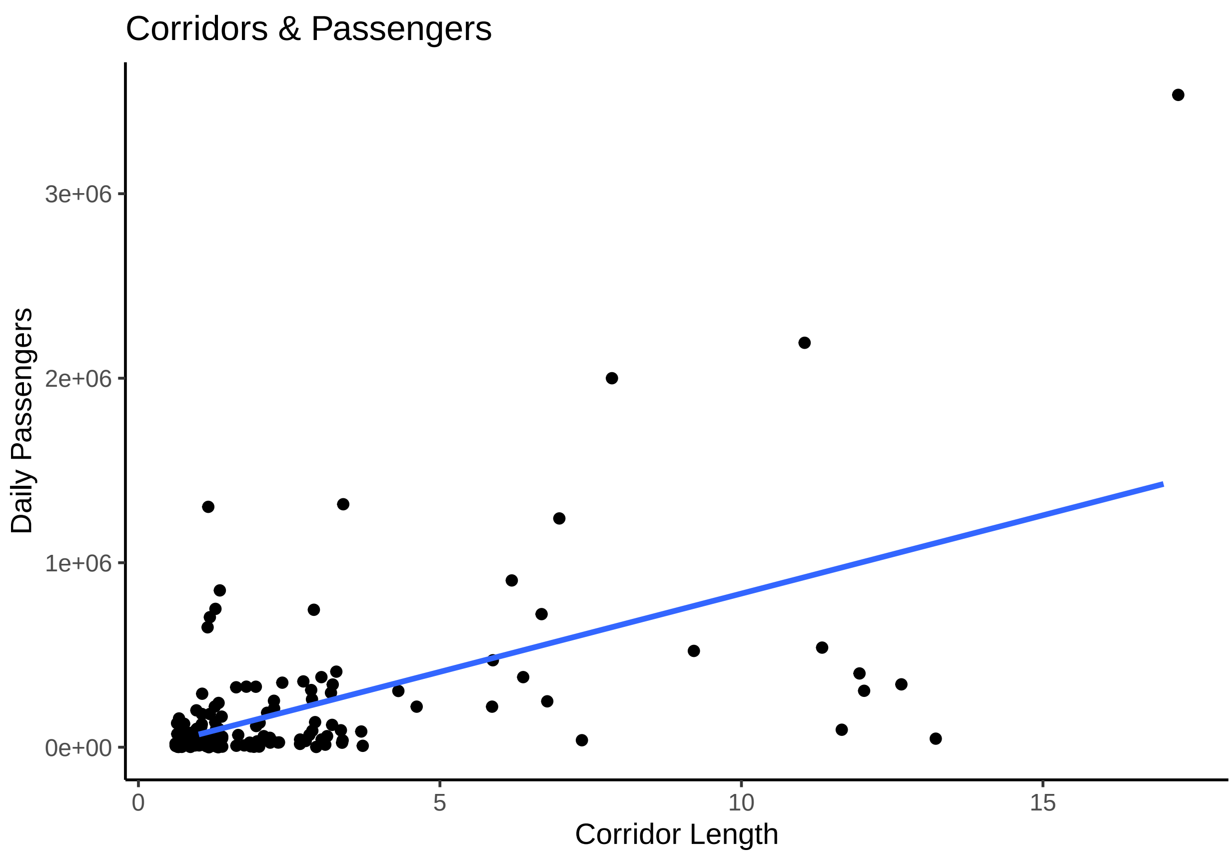

```
ggplot(kontinenten, aes(corridors)) +
geom_histogram(aes(y=..density..),
colour="black", fill="white") +
geom_density(alpha=0.6, fill="brown")
```

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




```
#Scatterplot to visualize bivariate correlation
ggplot(kontinenten, aes(corridors,passengers)) +
   geom_point(position='jitter') +
   geom_smooth(method="lm",se=FALSE) +
   labs(title="Corridors & Passengers",
      x="Corridor Length",
      y="Daily Passengers") +
   theme_classic()
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
#Statistical Correlation test
cor.test(kontinenten$passengers,kontinenten$corridors)
```

```
##
##  Pearson's product-moment correlation
##
## data:  kontinenten$passengers and kontinenten$corridors
## t = 9.0822, df = 166, p-value = 3.052e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4653391 0.6691877
## sample estimates:
##       cor
## 0.5761549
```

Our Pearson R coefficient is 0.58 with a p-value of 0.0000000000000003052. Since our p-value is lower than the 0.05 statistical significance cut-off point, we can conclude that, there is a strong relationship between corridors and daily passengers so we can therefore reject the null hypothesis which says "There is no link between number of corridors and the number of daily passengers."

In conclusion, Latin America, especially Brazil, have produced the model BRT system in terms of functionality and patronage and Asia is close second. Africa, given the huge landmass could do way better in this area and hopefully, the situation improves soon through proper administration and efficient planning