

# DSCC/CSC/STAT 462 Assignment 4

Due November 16, 2023 by 11:59 p.m.

```
# population
population_df <- read.csv("D:/Stats Project/Billionaires Statistics Dataset.csv")

# random sample
set.seed(123)
sample_df <- population_df[sample(nrow(population_df), 1500), ]
# reset index
sample_df$index <- seq_len(nrow(sample_df))
rownames(sample_df) <- NULL
```

1. Is average final worth of inherited billionaires equal to the average final worth of self-made billionaires.

$$H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 \neq \mu_2 \quad \alpha = 0.05$$

```
split_df_1 <- split(sample_df, sample_df$selfMade)
split_df_1p <- split(population_df, population_df$selfMade)

x_1 <- c(split_df_1$True$finalWorth)
y_1 <- c(split_df_1$False$finalWorth)

t.test(x_1, y_1, paired=F, conf.level=0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  x_1 and y_1
## t = -1.4057, df = 733.82, p-value = 0.1602
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2247.064   371.824
## sample estimates:
## mean of x mean of y
##  4530.321  5467.941
```

```

mean_x_11 <- mean(split_df_1p$True$finalWorth)
mean_x_12 <- mean(split_df_1p$False$finalWorth)

cat("Population mean final worth of self-made billionaires is", mean_x_11)

## Population mean final worth of self-made billionaires is 4465.397

cat("\nPopulation mean final worth of inherited billionaires is", mean_x_12)

##
## Population mean final worth of inherited billionaires is 4970.411

cat("\ndifference in population mean is", mean_x_11 - mean_x_12)

##
## difference in population mean is -505.0133

```

Since  $p > \alpha$ , we fail to reject. We don't have sufficient evidence to conclude that the average final net worth of self made billionaires and inherited billionaires is significantly different. With 95% confidence interval we can say that our true difference in mean will lie between (-2247.064, 371.824). True difference in population means is -505.0133, which certainly lies in our confidence interval found above.

2. Is average age of billionaires in USA equal to the average age of billionaires in China.

$H_0 : \mu_1 = \mu_2$   $H_1 : \mu_1 \neq \mu_2$   $\alpha = 0.05$

```

split_df_2 <- split(sample_df, sample_df$country)
split_df_2p <- split(population_df, population_df$country)

x_2 <- na.omit(c(split_df_2$`United States`$age))
y_2 <- na.omit(c(split_df_2$China$age))

t.test(x_2, y_2, paired=F, conf.level=0.95)

##
## Welch Two Sample t-test
##

```

```
## data:  x_2 and y_2
## t = 9.0803, df = 717.11, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  6.003986 9.316452
## sample estimates:
## mean of x mean of y
##  66.78522  59.12500
```

```
mean_x_21 <- mean(split_df_2p$`United States`$age, na.rm = TRUE)
mean_x_22 <- mean(split_df_2p$China$age, na.rm = TRUE)

cat("Population mean age of billionaires in USA is", mean_x_21)
```

```
## Population mean age of billionaires in USA is 67.30596
```

```
cat("\nPopulation mean age of billionaires in China is", mean_x_22)
```

```
##
## Population mean age of billionaires in China is 58.69643
```

```
cat("\ndifference in population mean is", mean_x_21 - mean_x_22)
```

```
##
## difference in population mean is 8.609532
```

Since  $p \ll \alpha$ , we reject. We have sufficient evidence to conclude that the average age of billionaires in USA and billionaires in China is significantly different.

With 95% confidence interval we can say that our true difference in mean will lie between (6.001081, 9.318364).

True difference in population means is 8.609937, which certainly lies in our confidence interval found above.

3. Is the variance in the final net worth of self-made billionaires less than the variance in the final net worth of inherited billionaires.

$$H_0 : \sigma_1^2 \geq \sigma_2^2 \quad H_1 : \sigma_1^2 < \sigma_2^2 \quad \alpha = 0.05$$

```
split_df_3 <- split(sample_df, sample_df$selfMade)
split_df_3p <- split(population_df, population_df$selfMade)
```

```
x_3 <- c(split_df_3$True$finalWorth)
y_3 <- c(split_df_3$False$finalWorth)
```

```
var.test(x_3, y_3, alt="less", conf.level=0.95)
```

```
##
## F test to compare two variances
##
## data: x_3 and y_3
## F = 0.59033, num df = 1028, denom df = 470, p-value = 2.795e-12
## alternative hypothesis: true ratio of variances is less than 1
## 95 percent confidence interval:
## 0.0000000 0.6708215
## sample estimates:
## ratio of variances
## 0.5903325
```

```
var_x_31 <- var(split_df_3p$True$finalWorth)
var_x_32 <- var(split_df_3p$False$finalWorth)
```

```
cat("Population variance of final worth of self-made billionaires is", var_x_31)
```

```
## Population variance of final worth of self-made billionaires is 89402684
```

```
cat("\nPopulation variance of final worth of inherited billionaires is", var_x_32)
```

```
##
## Population variance of final worth of inherited billionaires is 112660877
```

```
cat("\nratio of variances is", var_x_31/var_x_32)
```

```
##
## ratio of variances is 0.7935557
```

Since  $p \ll \alpha$ , we reject. We have sufficient evidence to conclude that the variance in the final worth of self-made billionaires is less than the variance in final worth of inherited billionaires.

With 95% confidence interval we can say that our true ratio of variances will lie between

(0.0000000, 0.6708215).

True ratio of variances in population is 0.7935557, which lies outside our 95% confidence interval found above, which means that we need a more precise confidence interval(i.e, even lower  $\alpha$ ) to contain the true ratio of variances.

4.Inference for correlation on GDP of the billionaire's country and their final worth.

$H_0 : \rho = 0$   $H_1 : \rho \neq 0$   $\alpha = 0.05$

```
string_to_numeric <- function(x) as.numeric(gsub("[^0-9.]", "", x))
```

```
x_4_ <- sapply(c(sample_df$gdp_country), string_to_numeric)
```

```
y_4_ <- c(sample_df$finalWorth)
```

```
complete_cases_41 <- complete.cases(x_4_, y_4_)
```

```
x_4 <- x_4_[complete_cases_41]
```

```
y_4 <- y_4_[complete_cases_41]
```

```
cor.test(x_4,y_4)
```

```
##
```

```
## Pearson's product-moment correlation
```

```
##
```

```
## data: x_4 and y_4
```

```
## t = 0.83061, df = 1415, p-value = 0.4063
```

```
## alternative hypothesis: true correlation is not equal to 0
```

```
## 95 percent confidence interval:
```

```
## -0.03003413 0.07406551
```

```
## sample estimates:
```

```
## cor
```

```
## 0.02207552
```

```
x_41_ <- sapply(c(population_df$gdp_country), string_to_numeric)
```

```
y_41_ <- c(population_df$finalWorth)
```

```
complete_cases_42 <- complete.cases(x_41_, y_41_)
```

```
x_41 <- x_41_[complete_cases_42]
```

```
y_41 <- y_41_[complete_cases_42]
```

```
cat("Population correlation between GDP of Country and final worth is", cor(x_41,y_41))
```

```
## Population correlation between GDP of Country and final worth is 0.03758884
```

Since  $p > \alpha$ , we fail to reject. We don't have sufficient evidence to conclude that the GDP of Billionaire's country and their final worth are linearly correlated.

With 95% confidence interval we can say that our true correlation will lie between (-0.03003413, 0.07406551).

True correlation in population is 0.03758884, which certainly lies in our confidence interval found above.

5. Is proportion of self-made billionaires in eastern countries equal to the proportion of self-made billionaires in western countries.

$H_0 : p_1 = p_2$   $H_1 : p_1 \neq p_2$   $\alpha = 0.05$

```
split_df_5 <- split(sample_df, sample_df$continent)
split_df_5p <- split(population_df, population_df$continent)

n_eastern_5 = length(split_df_5$Asia$selfMade) + length(split_df_5$Africa$selfMade)
n_western_5 = length(split_df_5$Europe$selfMade) + length(split_df_5$`North America`$selfMade)

x_eastern_5 <- length(split_df_5$Africa$selfMade[split_df_5$Africa$selfMade == "True"])
x_western_5 <- length(split_df_5$Europe$selfMade[split_df_5$Europe$selfMade == "True"])

prop.test(c(x_eastern_5, x_western_5), c(n_eastern_5, n_western_5), conf.level = 0.95, c)

##
## 2-sample test for equality of proportions without continuity correction
##
## data:  c(x_eastern_5, x_western_5) out of c(n_eastern_5, n_western_5)
## X-squared = 34.396, df = 1, p-value = 4.497e-09
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.09625292 0.18829258
## sample estimates:
##   prop 1    prop 2
## 0.7681388 0.6258661
```

```
n_eastern_5p = length(split_df_5p$Asia$selfMade) + length(split_df_5p$Africa$selfMade)
n_western_5p = length(split_df_5p$Europe$selfMade) + length(split_df_5p$`North America`$selfMade)

x_eastern_5p <- length(split_df_5p$Africa$selfMade[split_df_5p$Africa$selfMade == "True"])
```

```
x_western_5p <- length(split_df_5p$Europe$selfMade[split_df_5p$Europe$selfMade == "True"])

p_eastern = x_eastern_5p/n_eastern_5p
p_western = x_western_5p/n_western_5p

cat("Difference of population proportions is ", p_eastern-p_western)
```

```
## Difference of population proportions is 0.1183536
```

Since  $p \ll \alpha$ , we reject. We have sufficient evidence to conclude that the proportions of self-made billionaires in eastern countries is significantly different from the proportion of self-made billionaires in western countries.

with 95% confidence interval, we can say that the difference in the population proportions will lie between (0.09625292 , 0.18829258)

True difference in population proportions is 0.1183536 which certainly lies in the confidence interval. Moreover we can see that the confidence interval and the hypothesis test are consistent with each other.(Not contradicting each other)

6. One-way ANOVA based on continent of origin, to check if the average final worth of the billionaires is equal in each continent.

```
library(ggplot2)

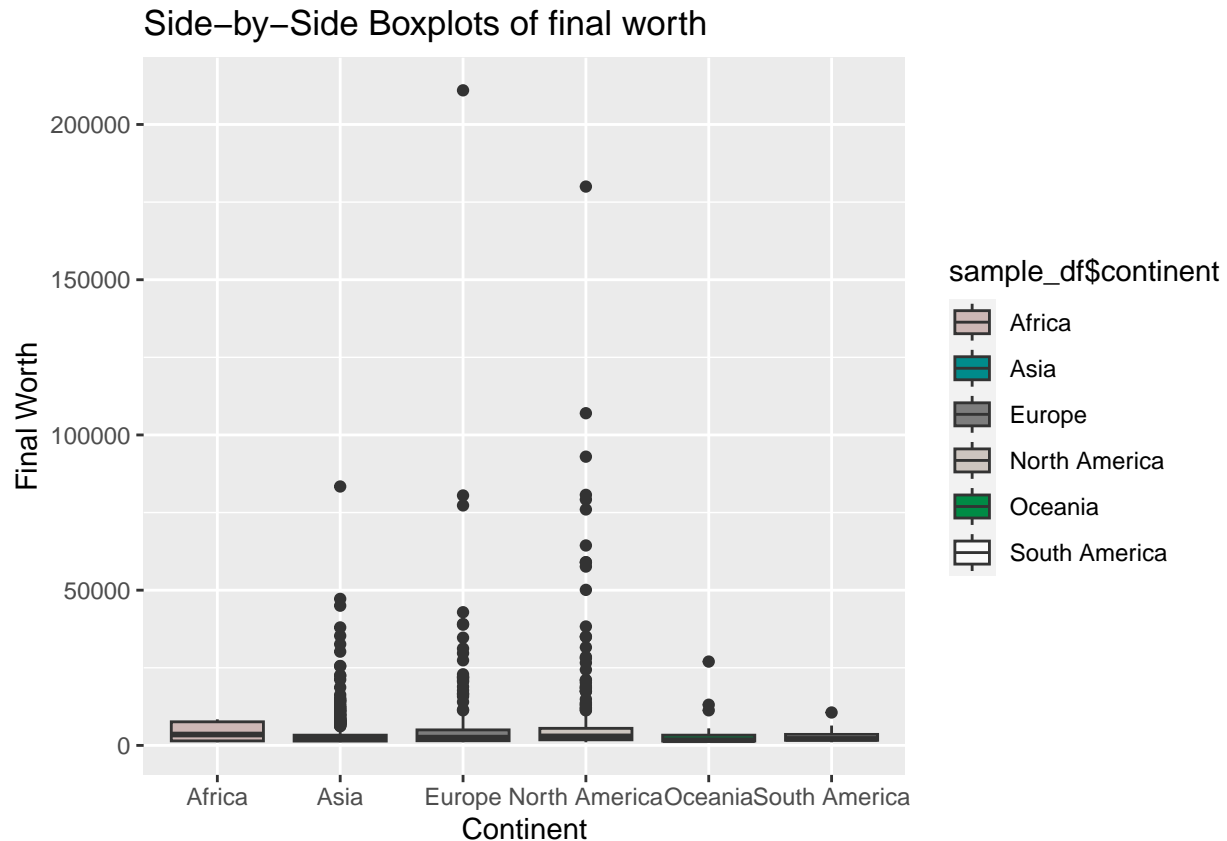
random_colors <- sample(colors(), 6)

# Create side-by-side boxplots with random colors using ggplot2
ggplot(sample_df, aes(x = sample_df$continent, y = sample_df$finalWorth, fill = sample_colors)) +
  geom_boxplot() +
  labs(title = "Side-by-Side Boxplots of final worth", x = "Continent", y = "Final Worth") +
  scale_fill_manual(values = random_colors)
```

```
## Warning: Use of 'sample_df$continent' is discouraged.
## i Use 'continent' instead.
```

```
## Warning: Use of 'sample_df$finalWorth' is discouraged.
## i Use 'finalWorth' instead.
```

```
## Warning: Use of 'sample_df$continent' is discouraged.
## i Use 'continent' instead.
```



From side-by-side boxplots of final worth, we can see that the variances of final worth for different continents are drastically different. So, we can't use parametric ANOVA. We will use Non-parametric ANOVA for this question.

$H_0 : \eta_1 = \eta_2 = \eta_3 = \dots = \eta_6$   $H_1 : H_0^c$   $\alpha = 0.05$

```
categories_6 <- sample_df$continent
values_6 <- sample_df$finalWorth

kruskal.test(values_6~categories_6, sample_df)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  values_6 by categories_6
## Kruskal-Wallis chi-squared = 57.486, df = 5, p-value = 4.016e-11
```

Since  $p \ll \alpha$ , we reject the null hypothesis. Now we use Dunn's test for post-hoc analysis. According to bonferroni correction,  $\alpha^* = \frac{\alpha}{\binom{6}{2}} = 0.00333$



```

library(dunn.test)
dunn.test(values_6, categories_6, method = "bonferroni")

## Kruskal-Wallis rank sum test
##
## data: values_6 and categories_6
## Kruskal-Wallis chi-squared = 57.4856, df = 5, p-value = 0
##
##
## Comparison of values_6 by categories_6
## (Bonferroni)
## Col Mean-|
## Row Mean | Africa Asia Europe North Am Oceania
## -----+-----
## Asia | 1.039736
## | 1.0000
## |
## Europe | 0.346446 -3.648319
## | 1.0000 0.0020*
## |
## North Am | -0.229786 -7.403688 -2.888914
## | 1.0000 0.0000* 0.0290
## |
## Oceania | 0.995316 0.168448 1.352879 2.368960
## | 1.0000 1.0000 1.0000 0.1338
## |
## South Am | 0.677871 -0.562746 0.778816 1.915113 -0.510803
## | 1.0000 1.0000 1.0000 0.4161 1.0000
##
## alpha = 0.05
## Reject Ho if p <= alpha/2

```

The pairwise comparisons of Asia-Europe and Asia-North America are the ones where  $p < \alpha$ , therefore we reject the null hypothesis for those pairs. We have sufficient evidence to say that

- i) The median final worth of billionaires in Asia and the median final worth of the billionaires in Europe are significantly different.
- ii) The median final worth of billionaires in Asia and the median final worth of the billionaires in North America are significantly different.

7. One-way ANOVA based on the continent of origin to check if the average age is the same in all continents or not.

```
library(ggplot2)

random_colors <- sample(colors(), 18)

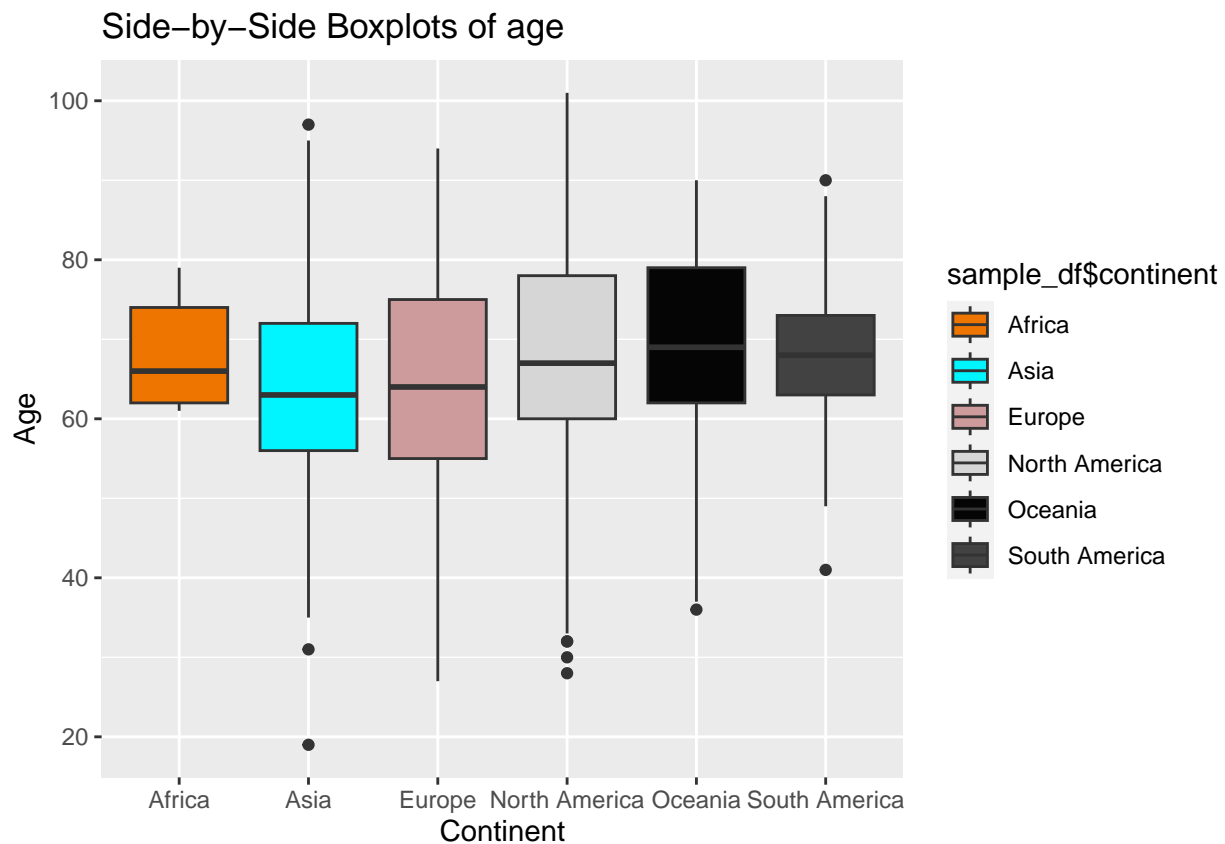
# Create side-by-side boxplots with random colors using ggplot2
ggplot(sample_df, aes(x = sample_df$continent, y = sample_df$age, fill = sample_df$continent)) +
  geom_boxplot() +
  labs(title = "Side-by-Side Boxplots of age", x = "Continent", y = "Age") +
  scale_fill_manual(values = random_colors)

## Warning: Use of 'sample_df$continent' is discouraged.
## i Use 'continent' instead.

## Warning: Use of 'sample_df$age' is discouraged.
## i Use 'age' instead.

## Warning: Use of 'sample_df$continent' is discouraged.
## i Use 'continent' instead.

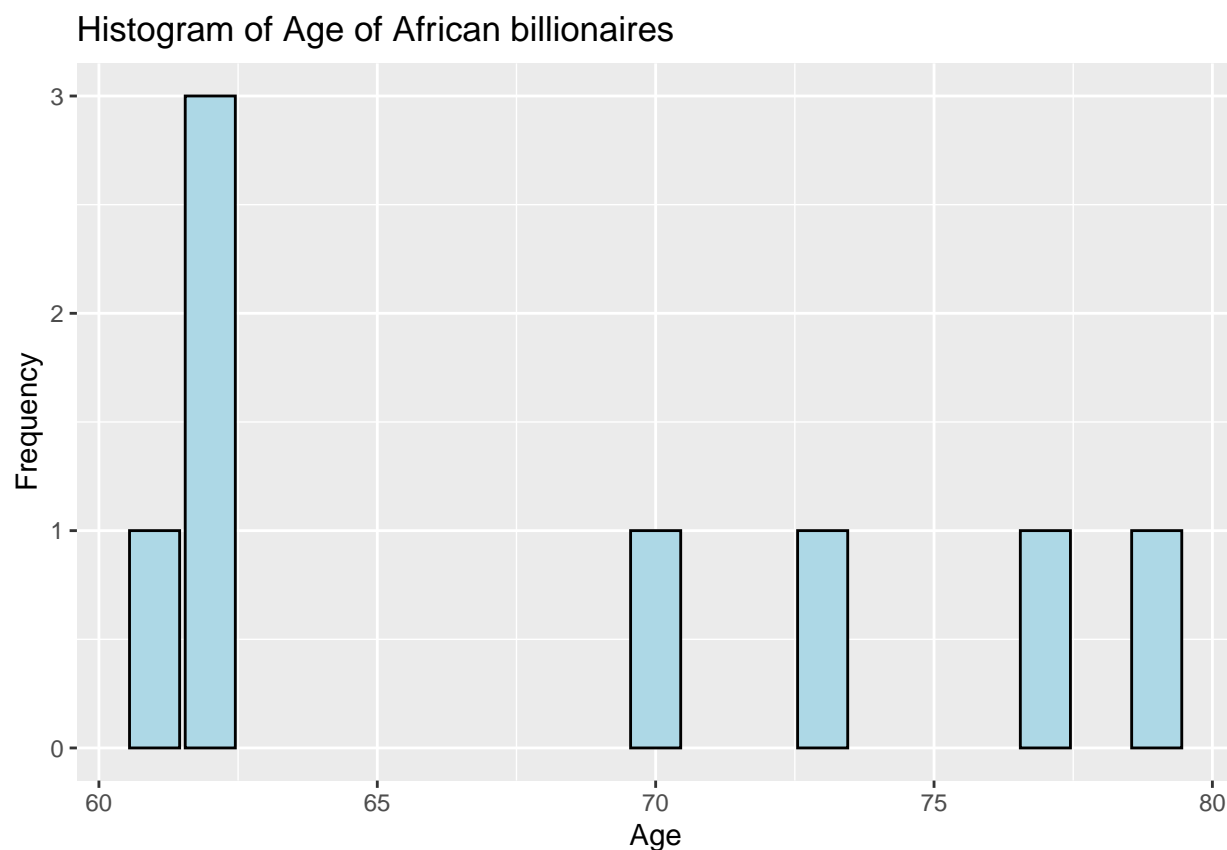
## Warning: Removed 36 rows containing non-finite values ('stat_boxplot()').
```



The equal variance assumptions looks reasonable here.

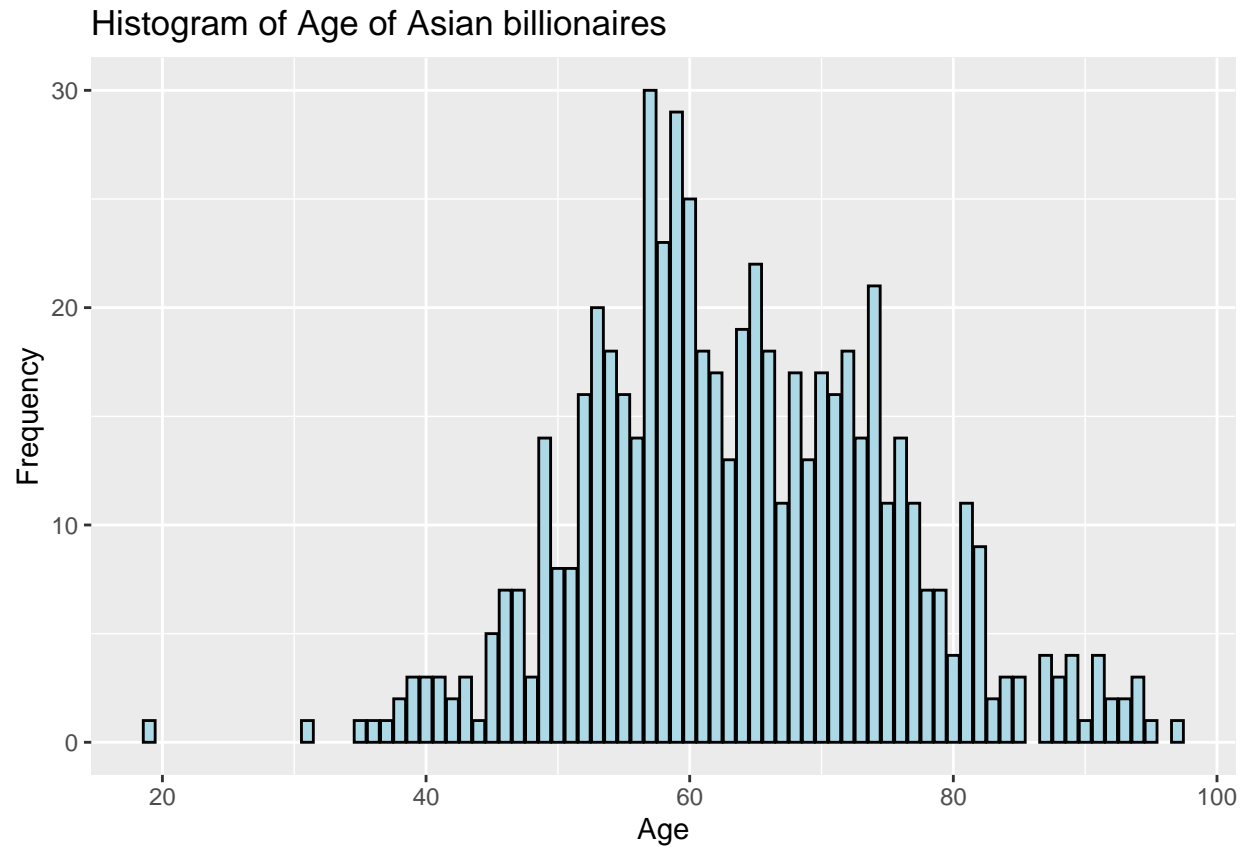
```
split_df_7 <- split(sample_df, sample_df$continent)

africa_7 <- split_df_7$Africa$Age
ggplot(data = data.frame(Category = africa_7), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of African billionaires", x = "Age", y = "Frequency")
```



```
asia_7 <- split_df_7$Asia$Age
ggplot(data = data.frame(Category = asia_7), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of Asian billionaires", x = "Age", y = "Frequency")
```

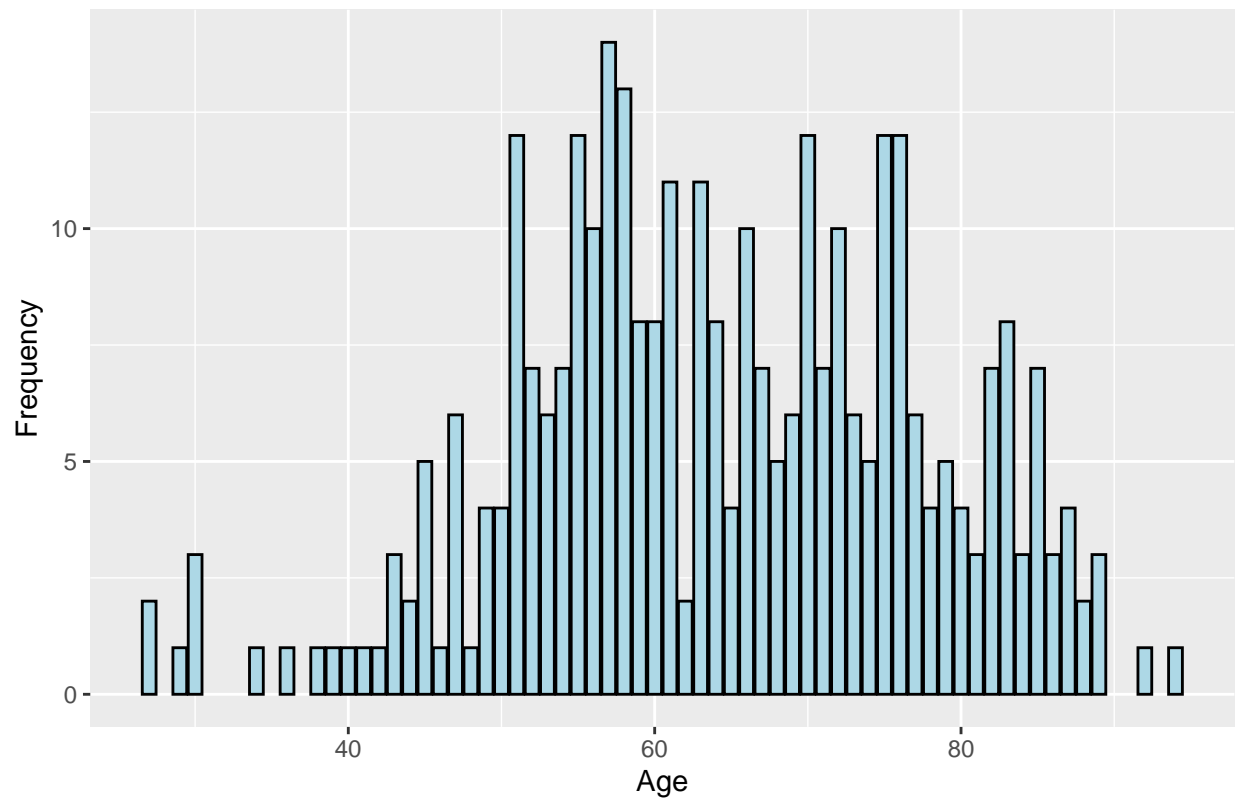
## Warning: Removed 20 rows containing non-finite values ('stat\_count()').



```
Europe_7 <- split_df_7$Europe$Age
ggplot(data = data.frame(Category = Europe_7), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of European billionaires", x = "Age", y = "Frequency")
```

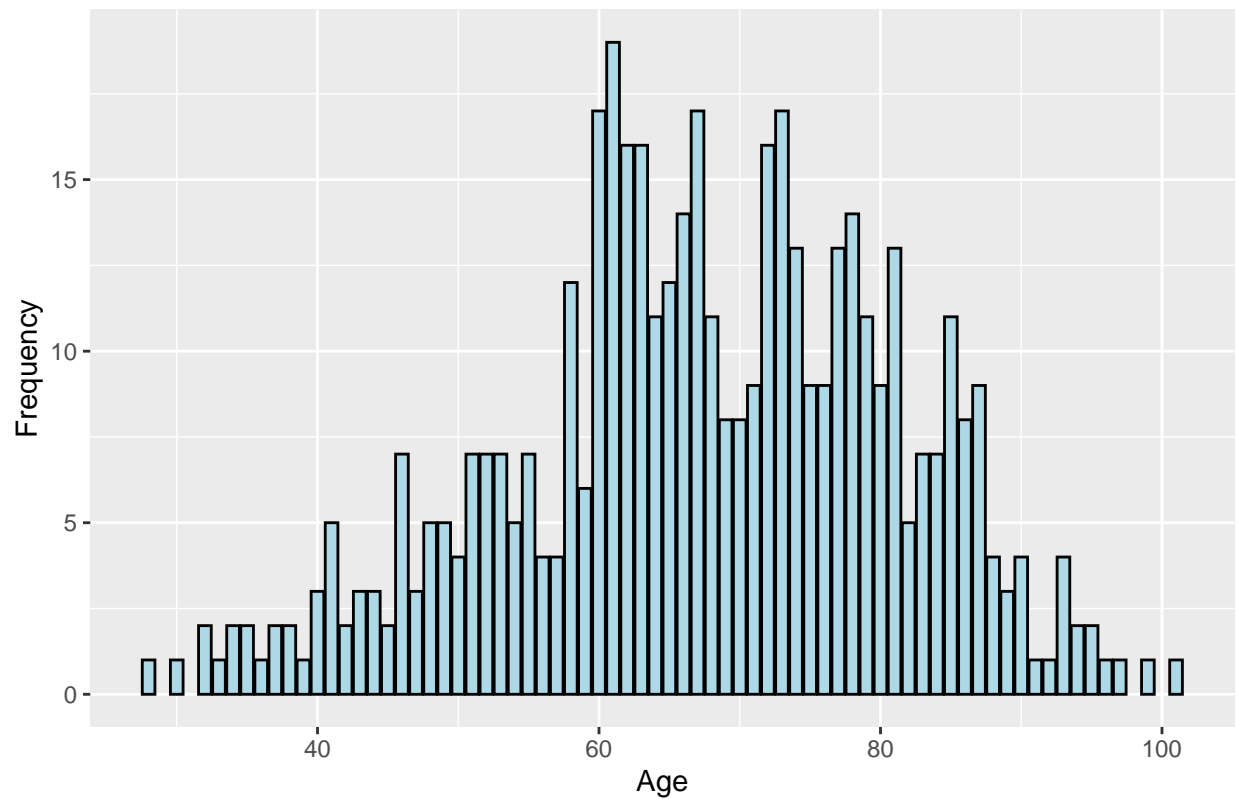
```
## Warning: Removed 14 rows containing non-finite values ('stat_count()').
```

Histogram of Age of European billionaires

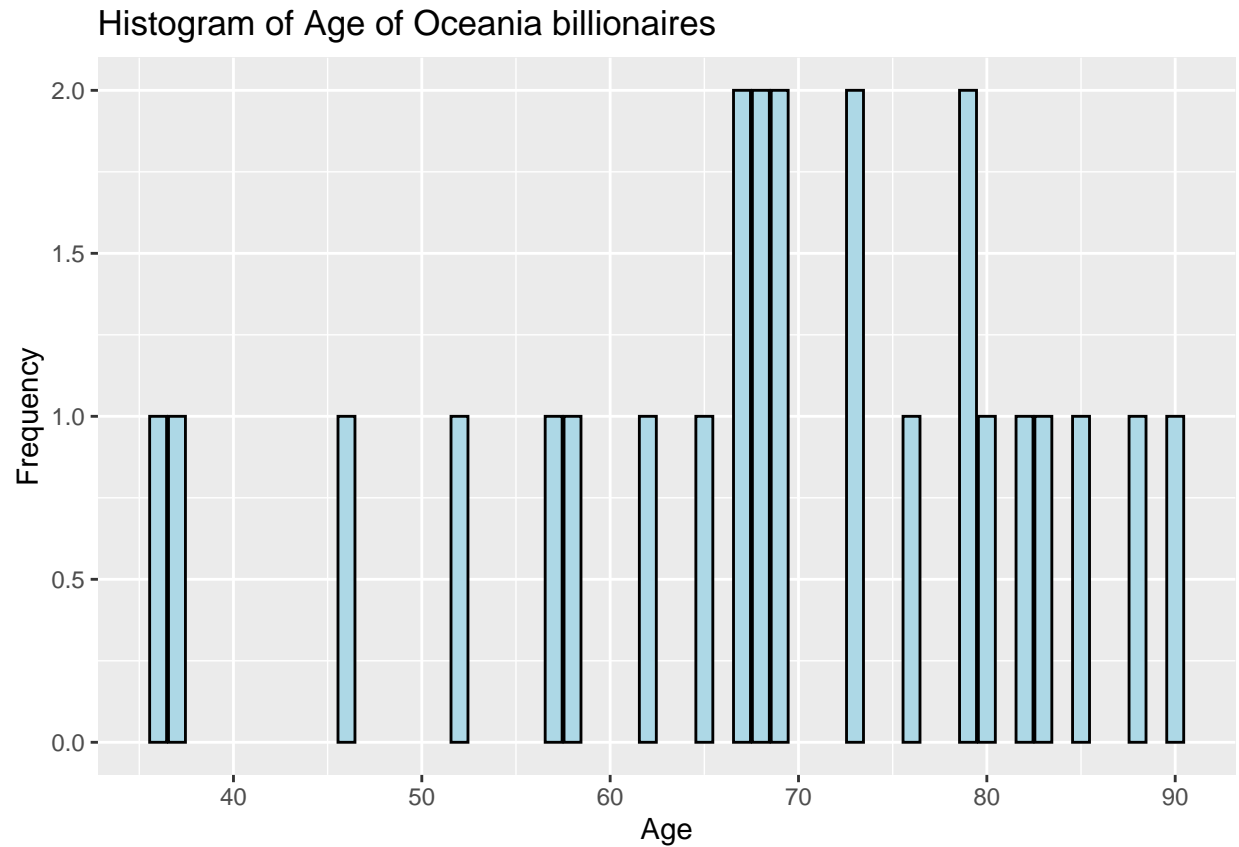


```
NorthAmerica_7 <- split_df_7$`North America`$age
ggplot(data = data.frame(Category = NorthAmerica_7), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of North American billionaires", x = "Age", y = "Frequency")
```

Histogram of Age of North American billionaires

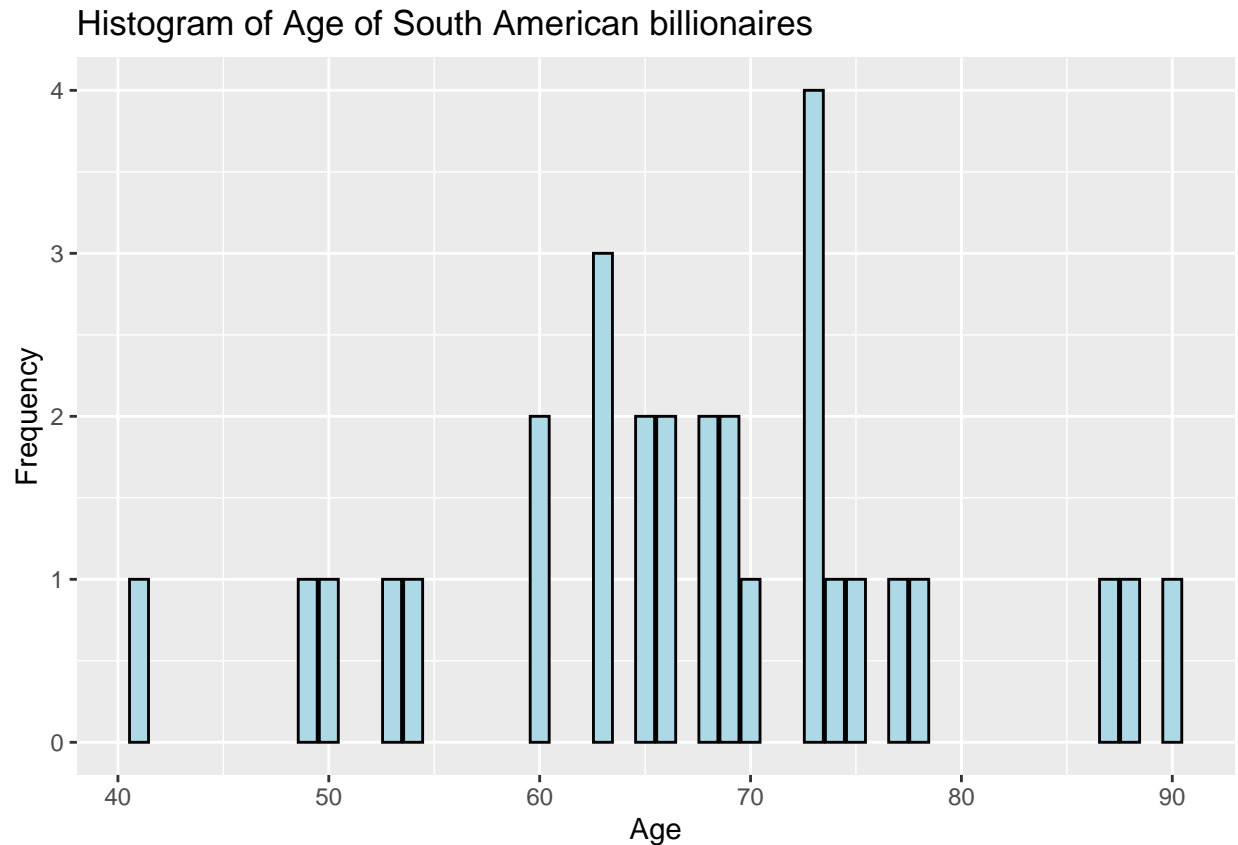


```
Oceania_7 <- split_df_7$Oceania$Age
ggplot(data = data.frame(Category = Oceania_7), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of Oceania billionaires", x = "Age", y = "Frequency")
```



```
SouthAmerica_7 <- split_df_7$`South America`$age
ggplot(data = data.frame(Category = SouthAmerica_7), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of South American billionaires", x = "Age", y = "Frequency")
```

```
## Warning: Removed 2 rows containing non-finite values ('stat_count()').
```



Each group follows a reasonably normal distribution as well. The observations in each group are independent as well.

$H_0 : \mu_1 = \mu_2 = \dots = \mu_6$   $H_1 : H_0^c$   $\alpha = 0.05$

```
fit_7 <- aov(sample_df$age~sample_df$continent)
summary(fit_7)
```

```
##               Df Sum Sq Mean Sq F value    Pr(>F)
## sample_df$continent      5   3979    795.7      4.87 0.000201 ***
## Residuals             1458 238246    163.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 36 observations deleted due to missingness
```

Since  $p < \alpha$ , we reject the null hypothesis. We have sufficient evidence to conclude that at least one of the equalities in the null hypothesis doesn't hold. Now we use post-hoc analysis for further investigation with bonferroni correction.

$$\alpha^* = \frac{\alpha}{\binom{6}{2}} = 0.00333$$



```
pairwise.t.test(sample_df$age, sample_df$continent, p.adj = "bonferroni")
```

```
##
## Pairwise comparisons using t tests with pooled SD
##
## data: sample_df$age and sample_df$continent
##
##           Africa Asia Europe North America Oceania
## Asia          1.00000 -          -          -          -
## Europe         1.00000 1.00000 -          -          -
## North America 1.00000 0.00012 0.02257 -          -
## Oceania        1.00000 1.00000 1.00000 1.00000 -
## South America 1.00000 1.00000 1.00000 1.00000 1.00000
##
## P value adjustment method: bonferroni
```

From pairwise comparisons, we can see that  $p < \alpha$  only for the pair Asia-North America, which means that we have sufficient evidence to conclude that the average age of billionaires in Asia and the average age of billionaires in North America are significantly different.

8. One-way ANOVA based on industries to check if the average final worth of billionaires across industries are the same.

```
library(ggplot2)

random_colors <- sample(colors(), 18)

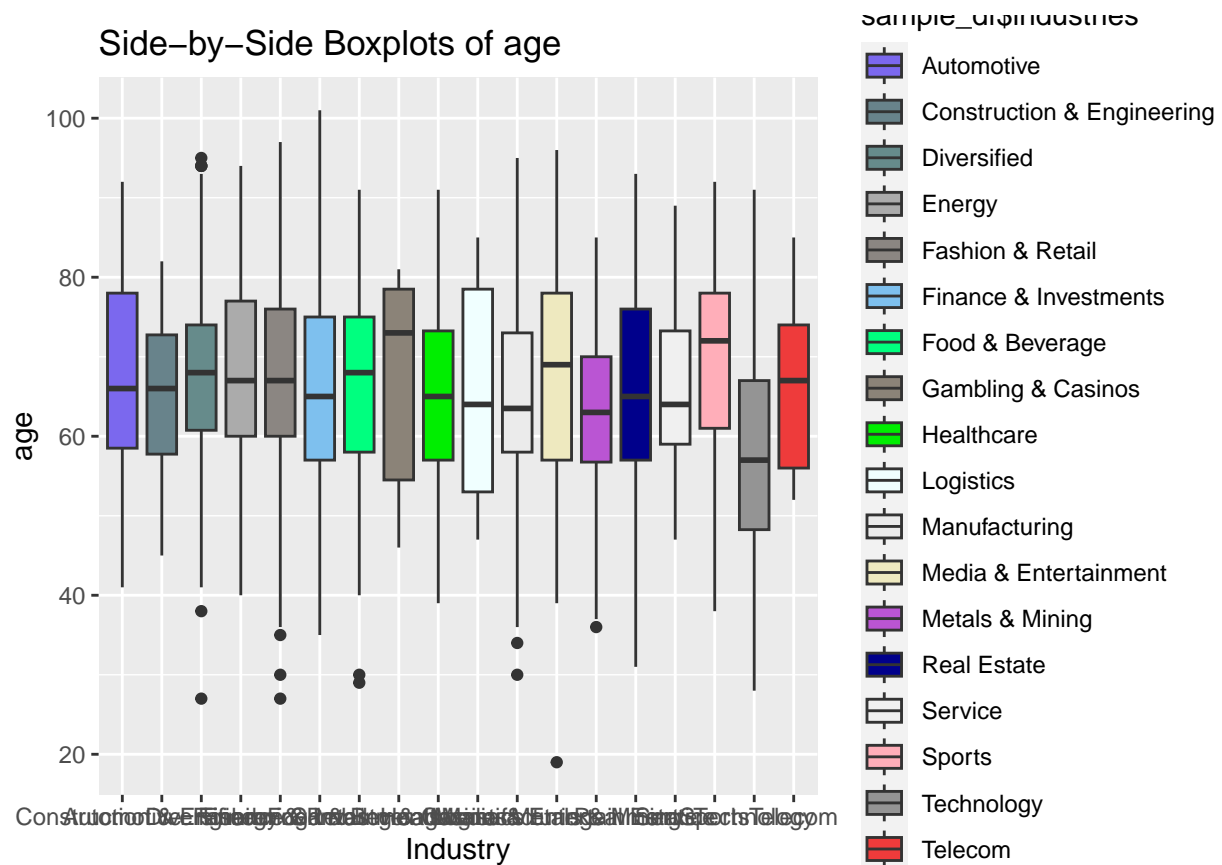
# Create side-by-side boxplots with random colors using ggplot2
ggplot(sample_df, aes(x = sample_df$industries, y = sample_df$age, fill = sample_df$industries)) +
  geom_boxplot() +
  labs(title = "Side-by-Side Boxplots of age", x = "Industry", y = "age") +
  scale_fill_manual(values = random_colors)
```

```
## Warning: Use of 'sample_df$industries' is discouraged.
## i Use 'industries' instead.
```

```
## Warning: Use of 'sample_df$age' is discouraged.
## i Use 'age' instead.
```

```
## Warning: Use of 'sample_df$industries' is discouraged.
## i Use 'industries' instead.
```

```
## Warning: Removed 36 rows containing non-finite values ('stat_boxplot()').
```



```
sorted_data_desc <- population_df[order(-population_df$finalWorth), ]

sorted_data_desc <- sorted_data_desc[-(1:533), ]

set.seed(1)
sample_df_1_5_billion <- sorted_data_desc[sample(nrow(sorted_data_desc), 1200), ]
# reset index
sample_df_1_5_billion$index <- seq_len(nrow(sample_df_1_5_billion))
rownames(sample_df_1_5_billion) <- NULL

random_colors <- sample(colors(), 40)

# Create side-by-side boxplots with random colors using ggplot2
ggplot(sample_df_1_5_billion, aes(x = sample_df_1_5_billion$industries, y = sample_df_1_5_billion$finalWorth)) +
  geom_boxplot() +
  labs(title = "Side-by-Side Boxplots of Final Worth", x = "Industry", y = "Final Worth") +
  scale_fill_manual(values = random_colors)
```

```
## Warning: Use of 'sample_df_1_5_billion$industries' is discouraged.
```

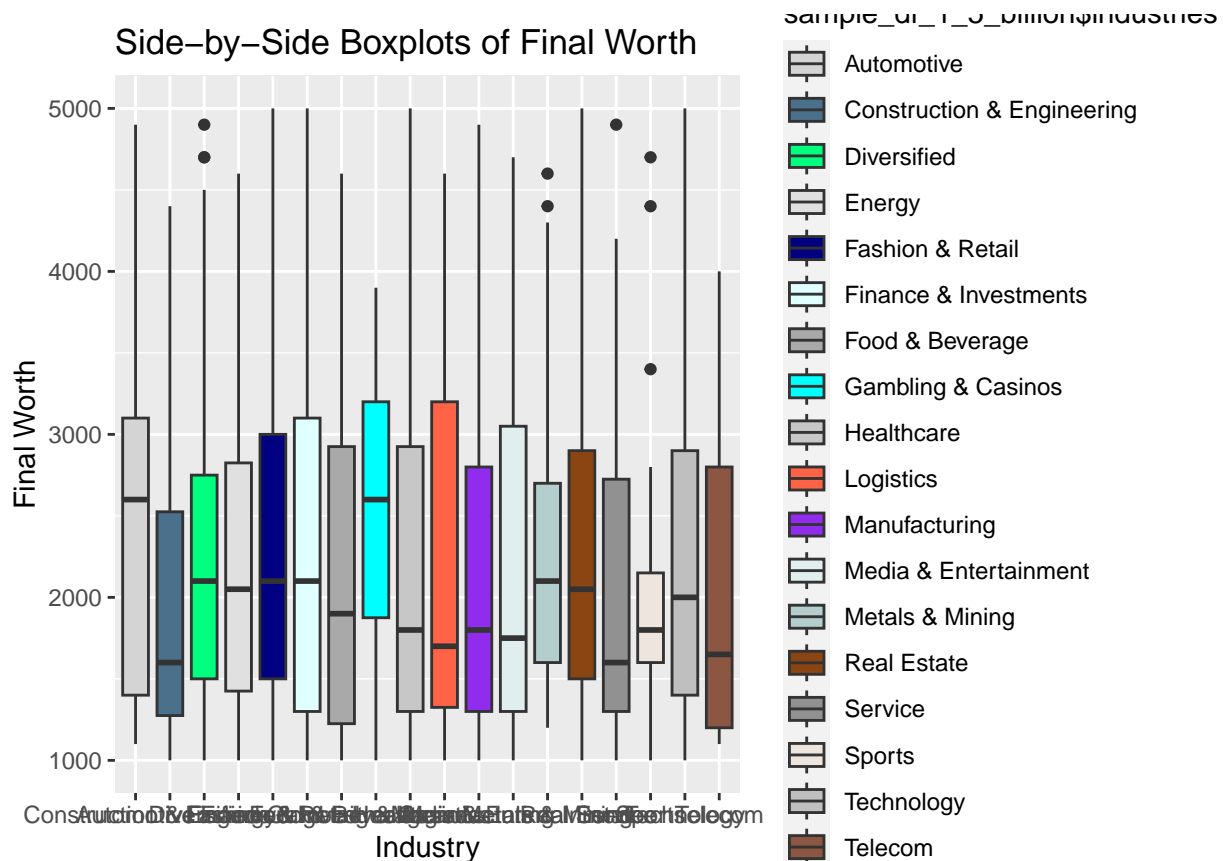
```
## i Use 'industries' instead.
```

```
## Warning: Use of 'sample_df_1_5_billion$finalWorth' is discouraged.
```

```
## i Use 'finalWorth' instead.
```

```
## Warning: Use of 'sample_df_1_5_billion$industries' is discouraged.
```

```
## i Use 'industries' instead.
```



```
# Create side-by-side boxplots with random colors using ggplot2
```

```
ggplot(sample_df_1_5_billion, aes(x = sample_df_1_5_billion$industries, y = sample_df_1_5_billion$finalWorth)) +  
  geom_boxplot() +  
  labs(title = "Side-by-Side Boxplots of age", x = "Industry", y = "age") +  
  scale_fill_manual(values = random_colors)
```

```
## Warning: Use of 'sample_df_1_5_billion$industries' is discouraged.
```

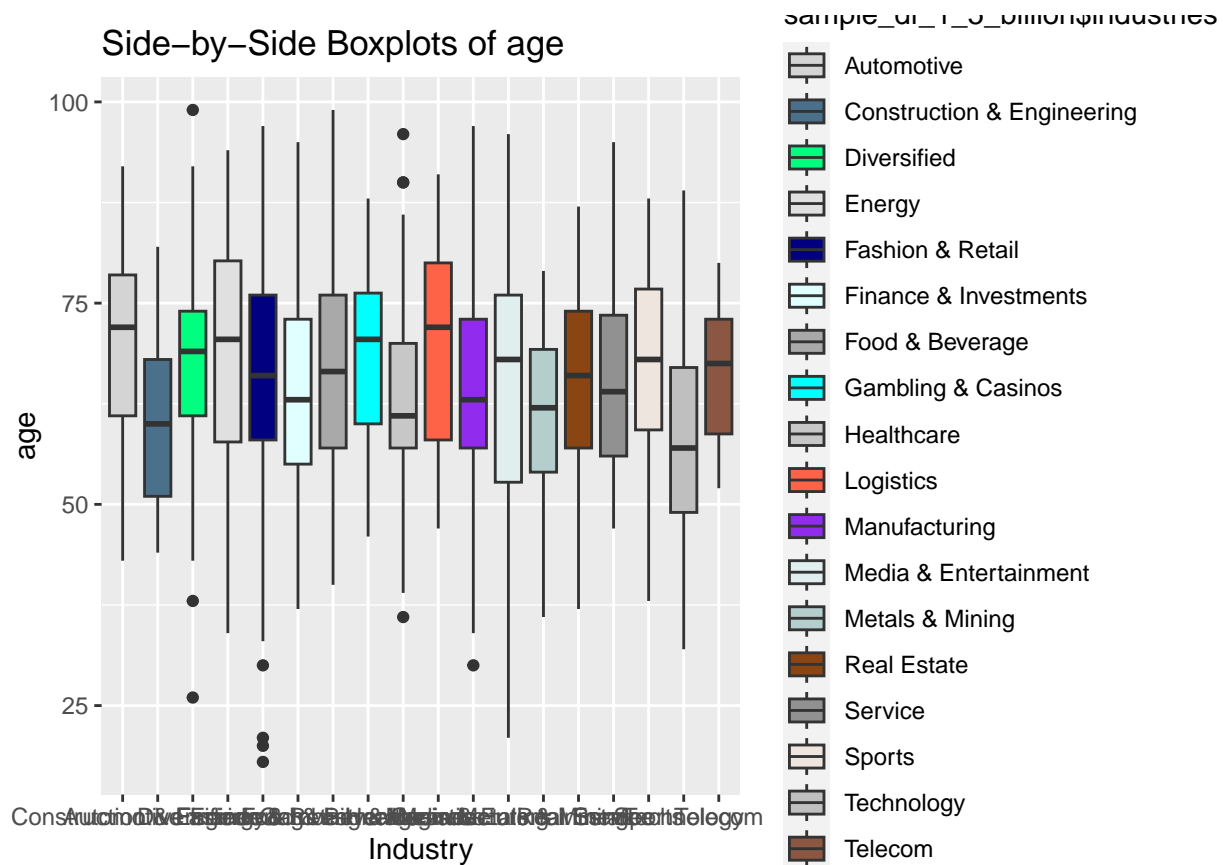
```
## i Use 'industries' instead.
```

```
## Warning: Use of 'sample_df_1_5_billion$age' is discouraged.
```

```
## i Use 'age' instead.
```

```
## Warning: Use of 'sample_df_1_5_billion$industries' is discouraged.
## i Use 'industries' instead.
```

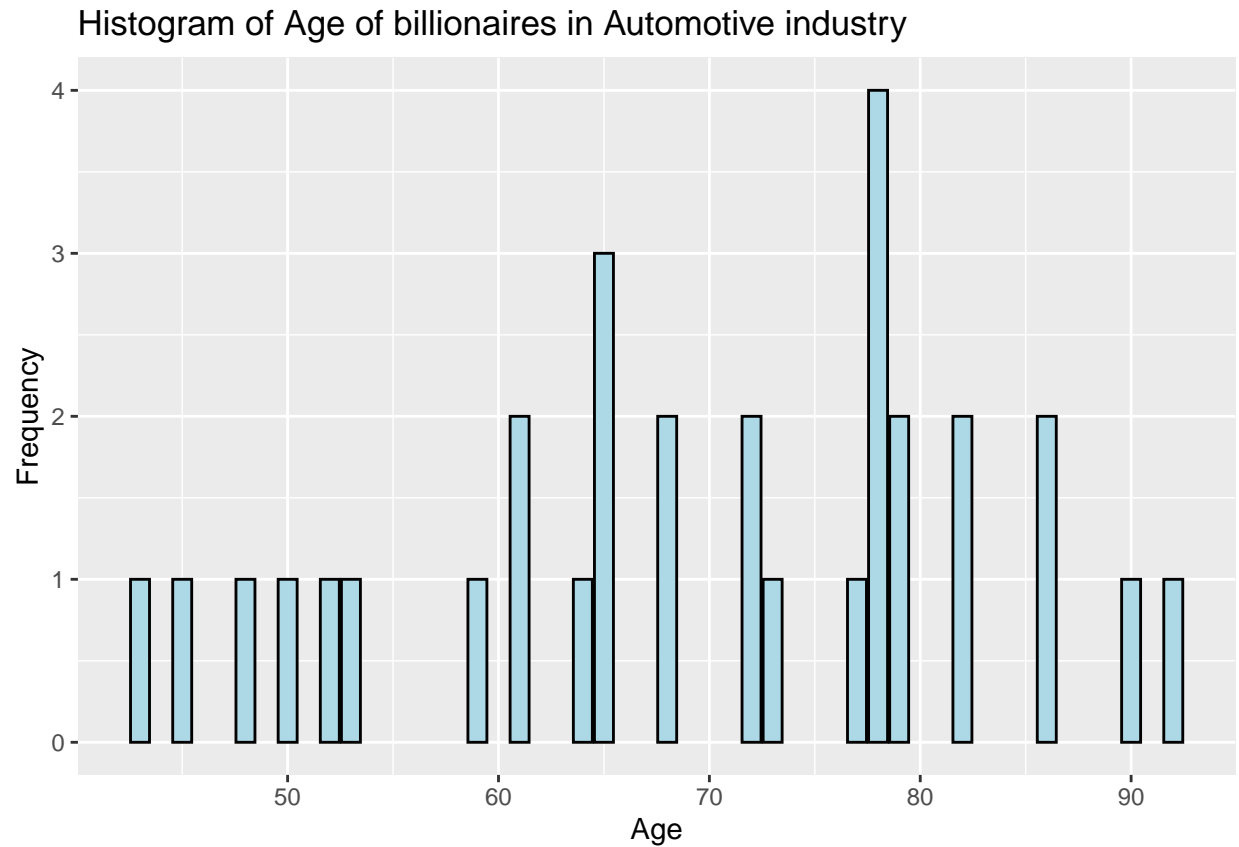
```
## Warning: Removed 37 rows containing non-finite values ('stat_boxplot()').
```



```
split_df_8 <- split(sample_df_1_5_billion, sample_df_1_5_billion$industries)
```

```
Automotive_8 <- split_df_8$Automotive$age
```

```
ggplot(data = data.frame(Category = Automotive_8), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of billionaires in Automotive industry", x = "Age", y = "Frequency")
```



```
Construction_Engineering_8 <- split_df_8$`Construction & Engineering`$age
ggplot(data = data.frame(Category = Construction_Engineering_8), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of billionaires in Construction & Engineering industry")
```



```
Diversified_8 <- split_df_8$Diversified$Age
ggplot(data = data.frame(Category = Diversified_8), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of billionaires in Diverse industry", x = "Age", y = "Frequency")
```

```
## Warning: Removed 2 rows containing non-finite values ('stat_count()').
```



```
Energy_8 <- split_df_8$Energy$Age
ggplot(data = data.frame(Category = Energy_8), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of billionaires in energy industry", x = "Age", y = "Fr
```

```
## Warning: Removed 2 rows containing non-finite values ('stat_count()').
```



```
Fashion_Retail_8 <- split_df_8$`Fashion & Retail`$age
ggplot(data = data.frame(Category = Fashion_Retail_8), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of billionaires in Fashion & Retail industry", x = "Age")
```

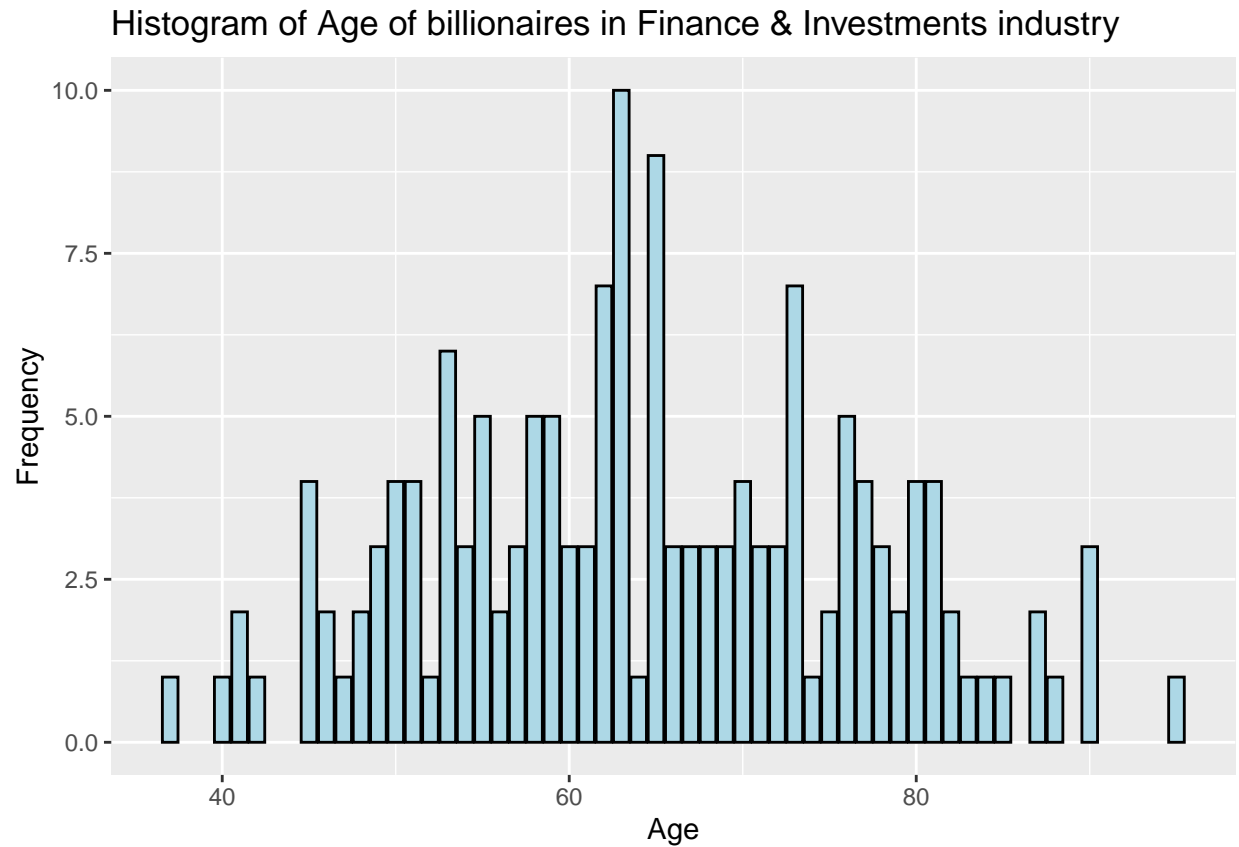
```
## Warning: Removed 4 rows containing non-finite values ('stat_count()').
```





```
Finance_Investments_8 <- split_df_8$`Finance & Investments`$age
ggplot(data = data.frame(Category = Finance_Investments_8), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of billionaires in Finance & Investments industry", x =
```

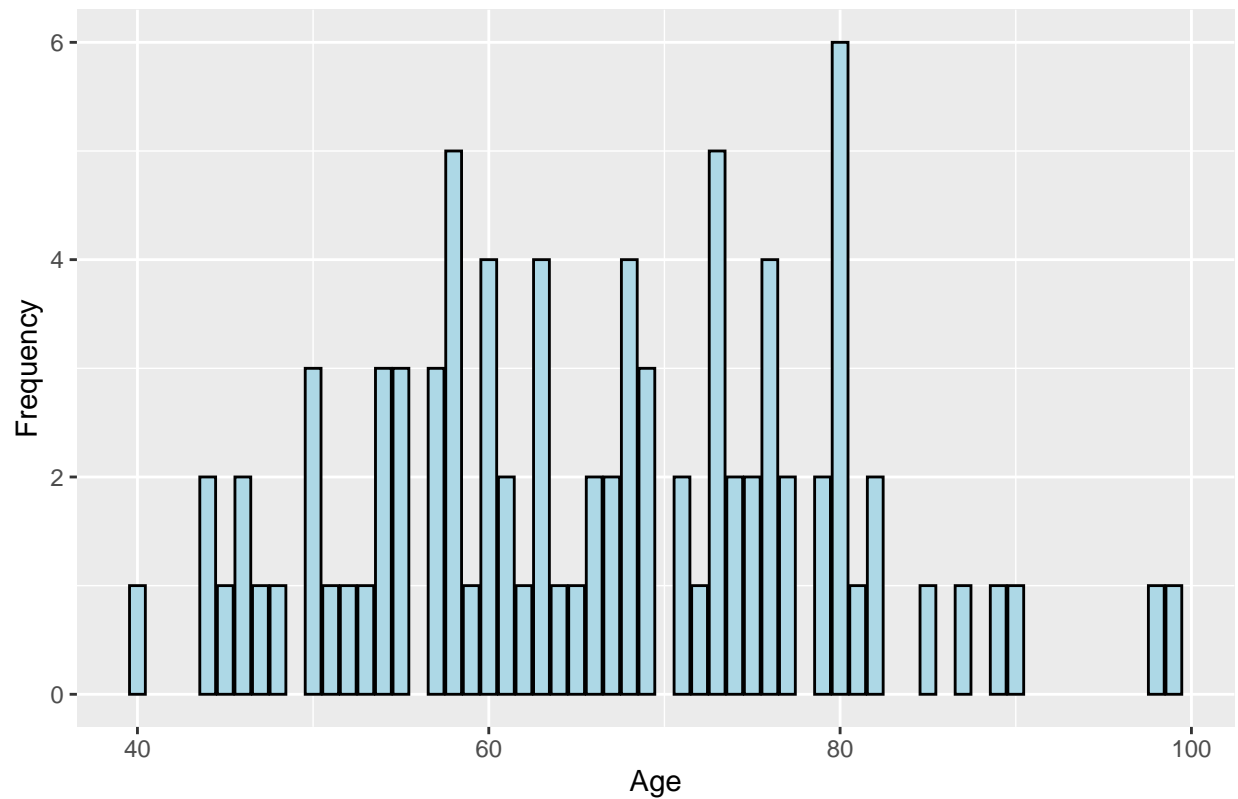
```
## Warning: Removed 3 rows containing non-finite values ('stat_count()').
```



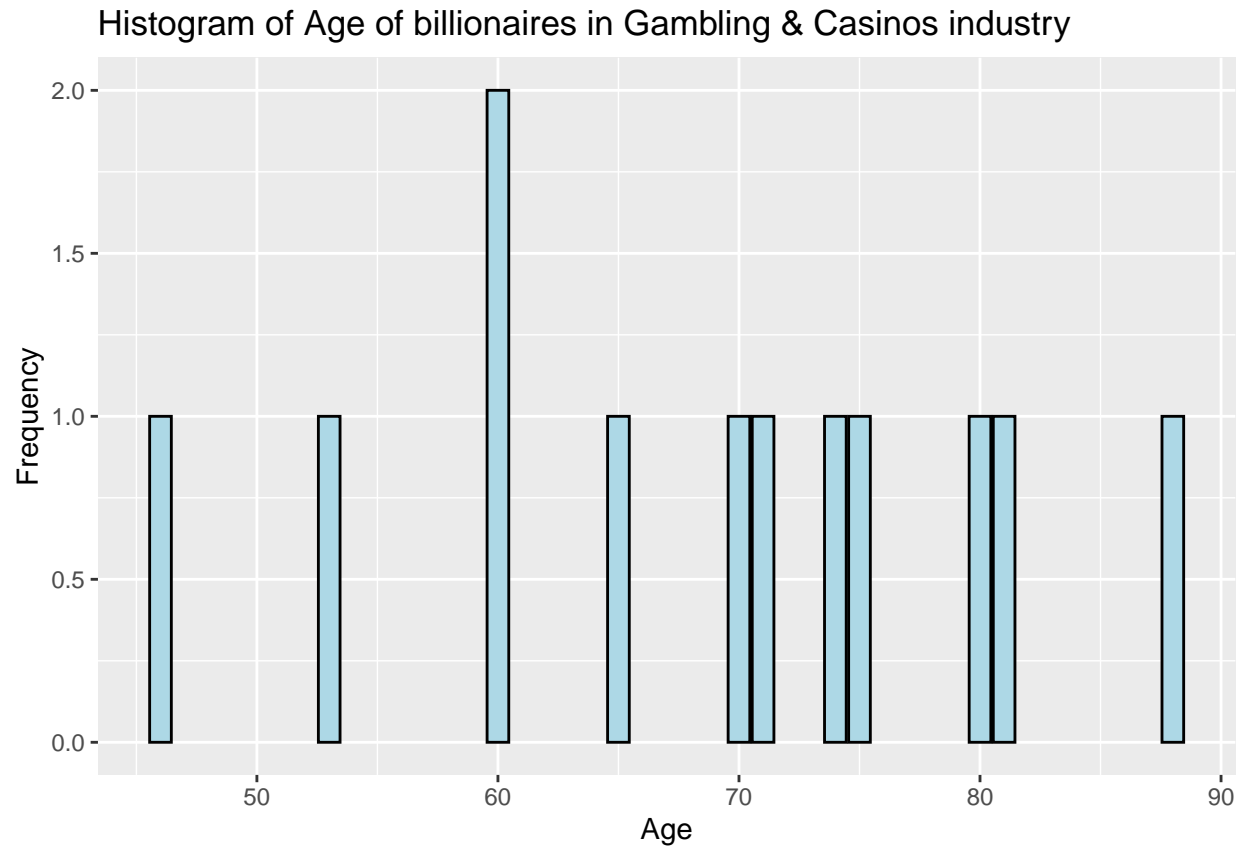
```
Food_Beverage_8 <- split_df_8$`Food & Beverage`$age
ggplot(data = data.frame(Category = Food_Beverage_8), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of billionaires in Food & Beverage industry", x = "Age")
```

```
## Warning: Removed 6 rows containing non-finite values ('stat_count()').
```

Histogram of Age of billionaires in Food & Beverage industry



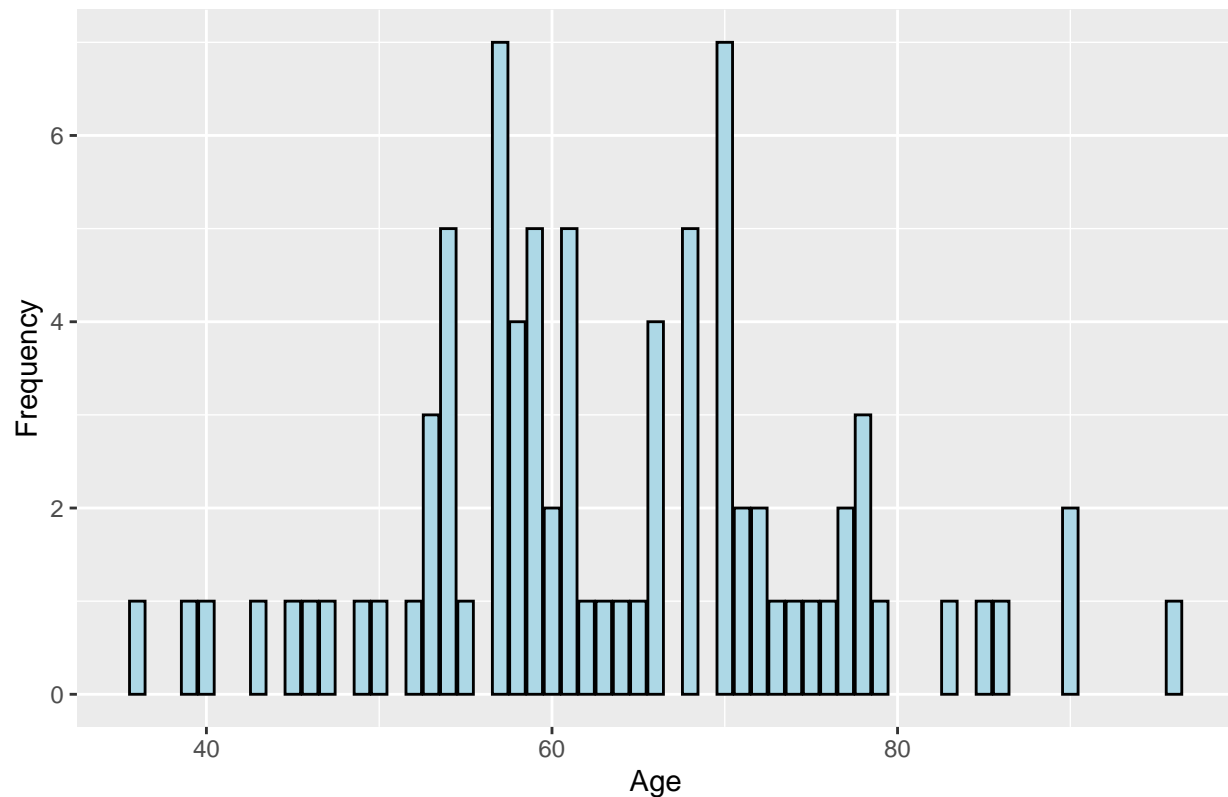
```
Gambling_Casinos_8 <- split_df_8$`Gambling & Casinos`$age
ggplot(data = data.frame(Category = Gambling_Casinos_8), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of billionaires in Gambling & Casinos industry", x = "A
```



```
Healthcare_8 <- split_df_8$Healthcare$Age
ggplot(data = data.frame(Category = Healthcare_8), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of billionaires in Healthcare industry", x = "Age", y =
```

```
## Warning: Removed 2 rows containing non-finite values ('stat_count()').
```

Histogram of Age of billionaires in Healthcare industry



```
Logistics_8 <- split_df_8$Logistics$Age
ggplot(data = data.frame(Category = Logistics_8), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of billionaires in Logistics industry", x = "Age", y =
```

```
## Warning: Removed 1 rows containing non-finite values ('stat_count()').
```



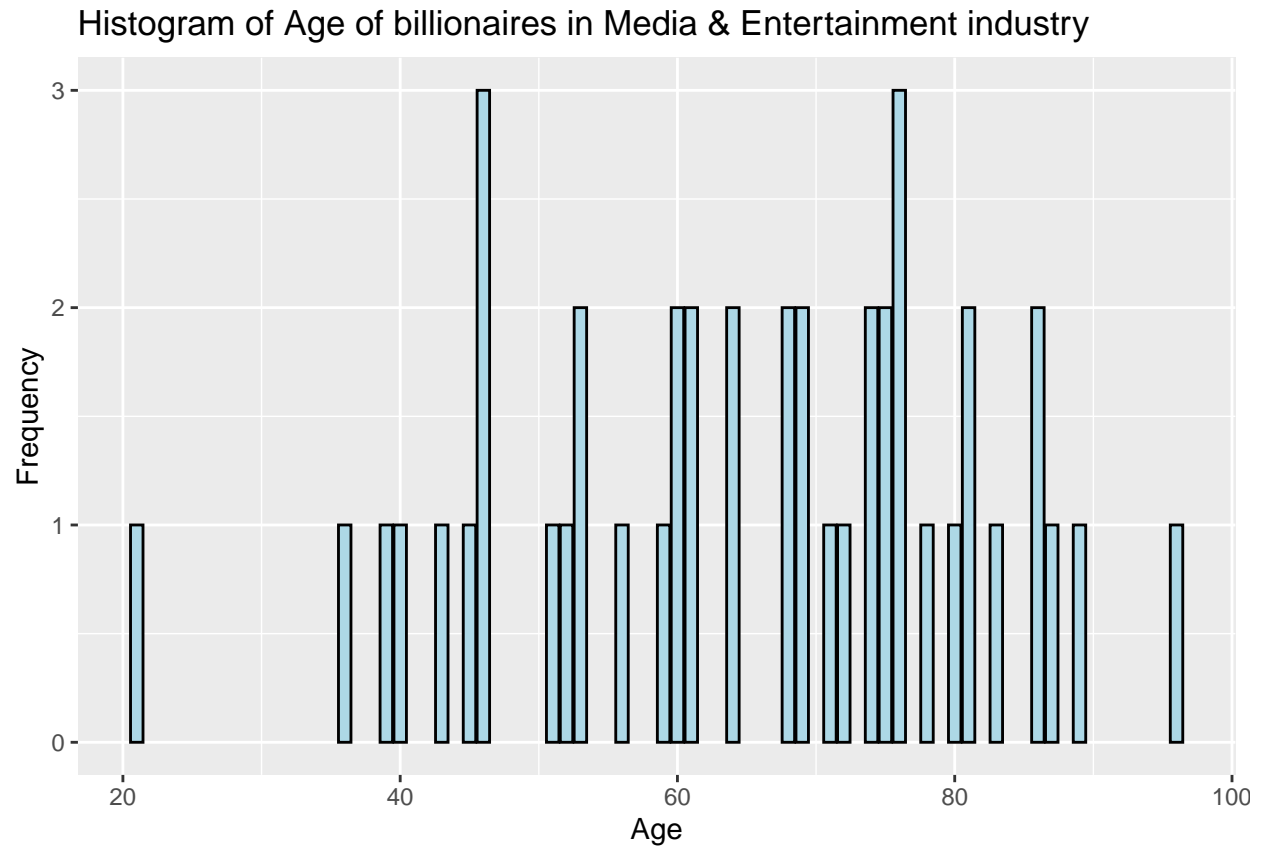
```
Manufacturing_8 <- split_df_8$Manufacturing$Age
ggplot(data = data.frame(Category = Manufacturing_8), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of billionaires in Manufacturing industry", x = "Age",
```

```
## Warning: Removed 6 rows containing non-finite values ('stat_count()').
```



```
Media_Entertainment_8 <- split_df_8$`Media & Entertainment`$age
ggplot(data = data.frame(Category = Media_Entertainment_8), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of billionaires in Media & Entertainment industry", x =
```

```
## Warning: Removed 2 rows containing non-finite values ('stat_count()').
```

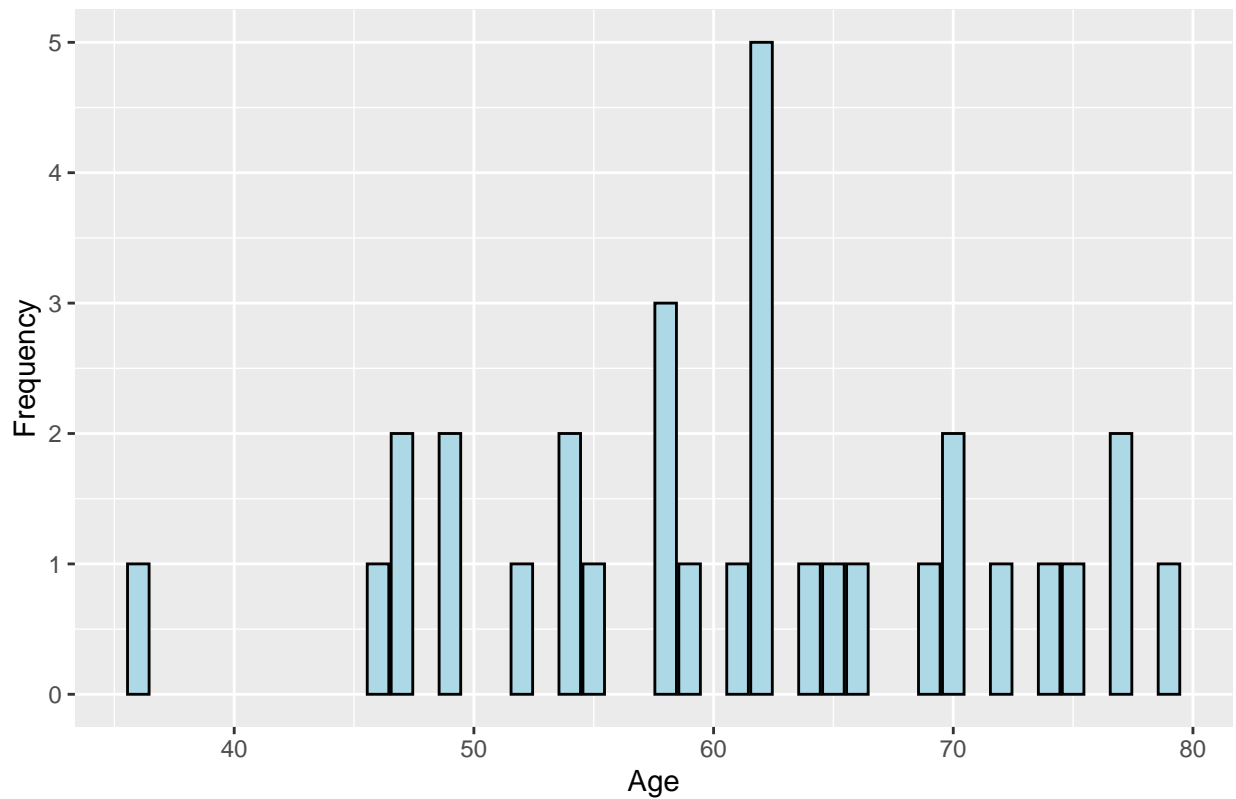


```
Metals_Mining_8 <- split_df_8$`Metals & Mining`$age
ggplot(data = data.frame(Category = Metals_Mining_8), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of billionaires in Metals & Mining industry", x = "Age")
```

```
## Warning: Removed 1 rows containing non-finite values ('stat_count()').
```



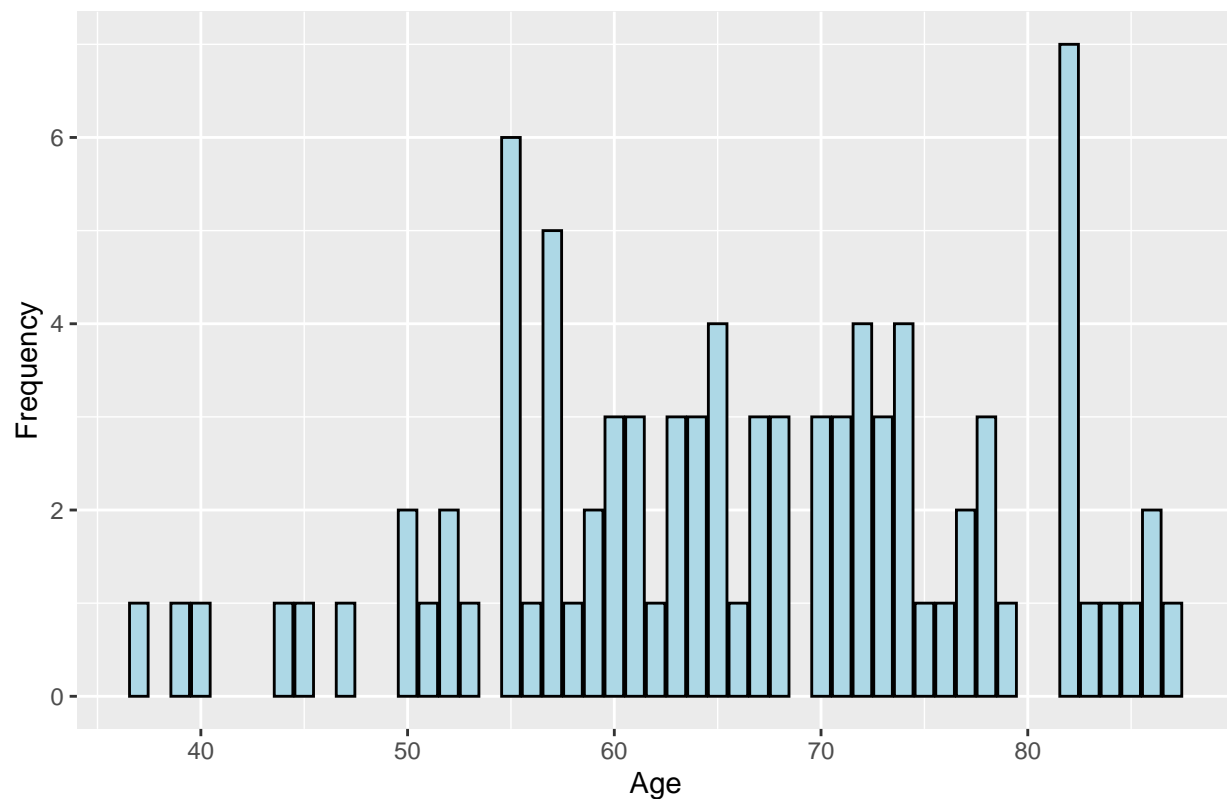
Histogram of Age of billionaires in Metals & Mining industry



```
Real_Estate_8 <- split_df_8$`Real Estate`$age
ggplot(data = data.frame(Category = Real_Estate_8), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of billionaires in Real Estate industry", x = "Age", y
```

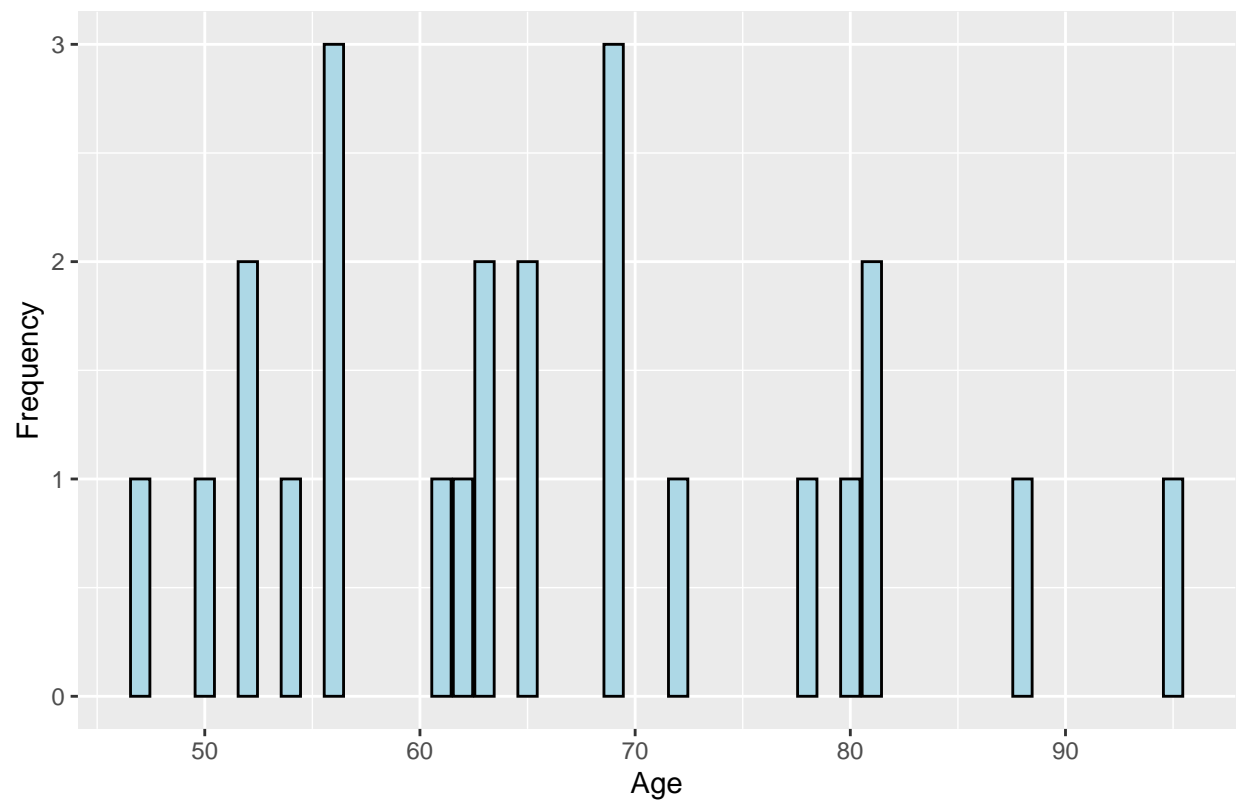
```
## Warning: Removed 1 rows containing non-finite values ('stat_count()').
```

Histogram of Age of billionaires in Real Estate industry

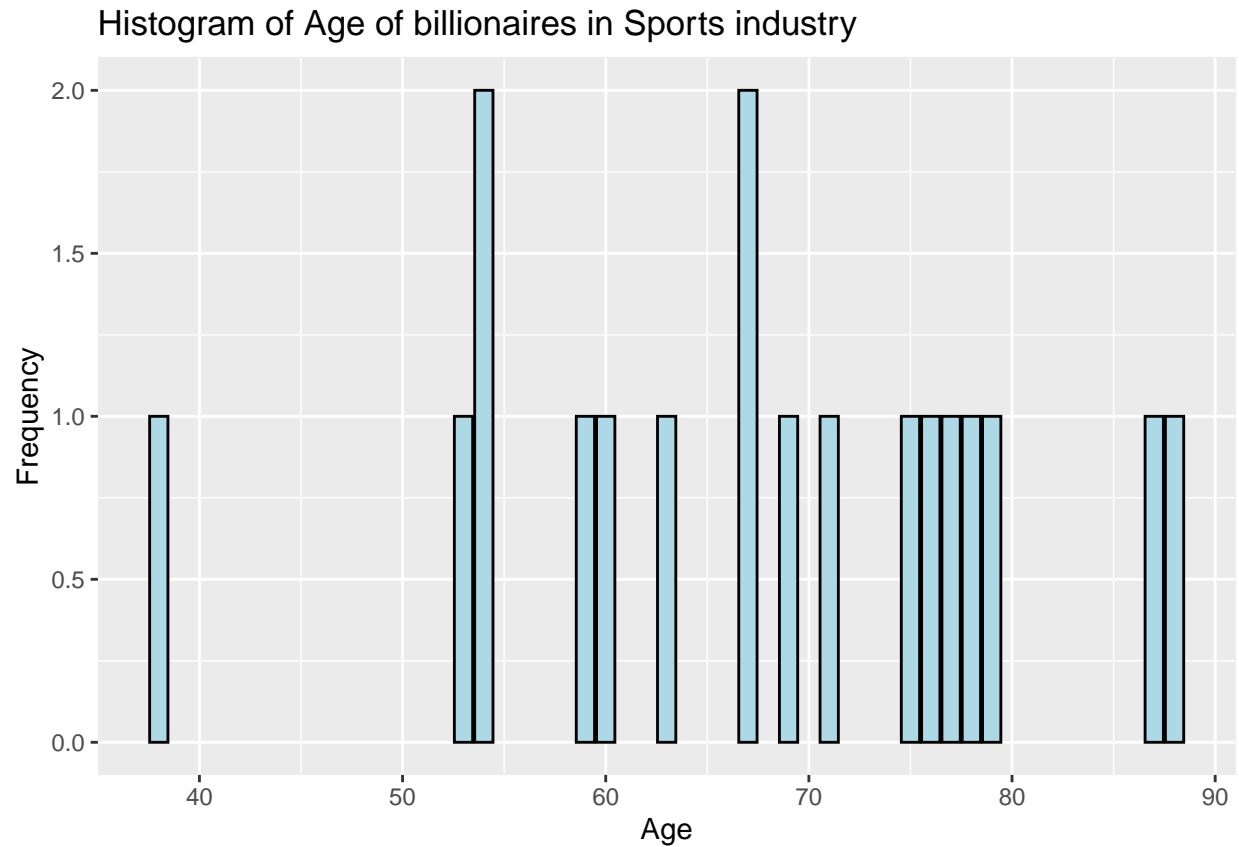


```
Service_8 <- split_df_8$Service$Age
ggplot(data = data.frame(Category = Service_8), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of billionaires in Service industry", x = "Age", y = "Frequency")
```

Histogram of Age of billionaires in Service industry



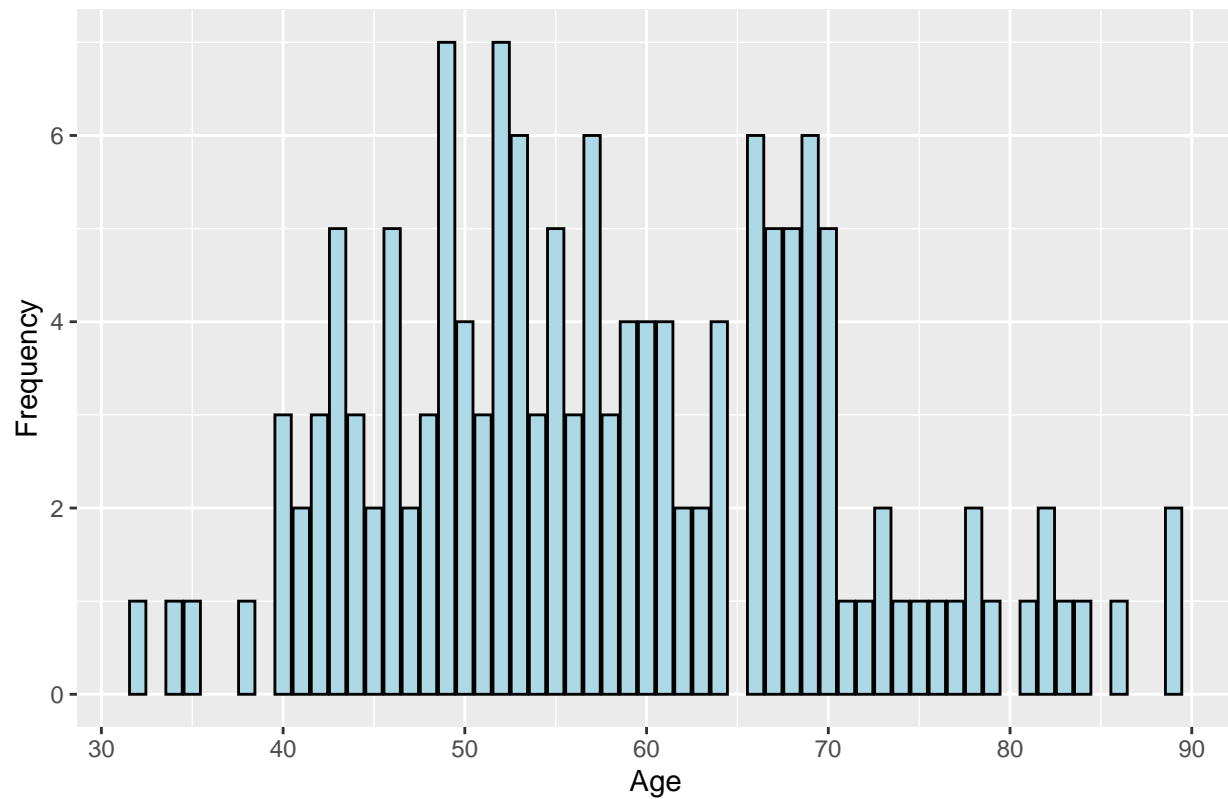
```
Sports_8 <- split_df_8$Sports$Age
ggplot(data = data.frame(Category = Sports_8), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of billionaires in Sports industry", x = "Age", y = "Fr
```



```
Technology_8 <- split_df_8$Technology$Age
ggplot(data = data.frame(Category = Technology_8), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of billionaires in Technology industry", x = "Age", y =
```

```
## Warning: Removed 7 rows containing non-finite values ('stat_count()').
```

Histogram of Age of billionaires in Technology industry



```
Telecom_8 <- split_df_8$Telecom$Age
ggplot(data = data.frame(Category = Telecom_8), aes(x = Category)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Age of billionaires in Telecom industry", x = "Age", y = "Frequency")
```

Histogram of Age of billionaires in Telecom industry

