---

title: "python to R Analysis"

author: "17F-8075"

output: pdf

```{r}

library(tidyverse)

library(naniar)

library(visdat)

bill_df <- read.csv("C:/Users/Talha/Downloads/forbes_billionaires.csv", header = TRUE, na.strings = c("NA","N/A",""))

head(bill_df)

glimpse(bill_df)

sum(duplicated(bill_df))

bill_df %>%

  count(Name) %>%

  filter (n > 1)

sum(is.na(bill_df))

miss_var_summary(bill_df)

bill_df %>%

  arrange(NetWorth) %>%

  vis_miss()
```

```

```{r}
bill_df_cleaned <- na.omit(bill_df) %>%
  distinct(Name, .keep_all = TRUE)

sum(is.na(bill_df_cleaned))

bill_df_cleaned %>%
  count(Name) %>%
  filter (n > 1)

print(bill_df$NetWorth)

bill_df_cleaned %>%
  ggplot(aes(x = NetWorth)) + geom_histogram(binwidth = 40, color = "blue", fill = "blue") + labs(title = "Net Worth")

```
```

```{r}

NetWorthLabel = c("1 - 88.5", "88.6 - 177")

bill_df_cleaned$NetWorth_Group = cut(bill_df_cleaned$NetWorth, breaks = c(1, 88.5, Inf), labels = NetWorthLabel, right = FALSE)

ggplot(data = bill_df_cleaned, mapping = aes(x = NetWorth_Group, y = Age)) + geom_boxplot(alpha = 0, color = "black", fill = "white") + geom_jitter(color = "red", alpha = 0.5) + labs(title = "Forbes bill_df 2021 by Age")

```