

# YZ\_BMC\_PARTII\_CODE

Yueran Zhang

2022-10-05

Now that you have created the .csv file, read it, calculate all the above statistics based on the generated cohort and write them in your report document.

Also, create a simple plot where X axis is the idu states and Y axis is percentage of infected people in each idu state.

```
library(ggplot2)
```

**Statistical check from my Cohort.csv file**

```
# Try to get the results only around with 3 decimal places
options(digits=3)
```

```
# Import dat
df <- read.csv("Cohort.csv", stringsAsFactors=FALSE)
```

```
# Check for the # of the population
population_size <- nrow(df)
population_size
```

```
## [1] 10000
```

```
#Statistical Results
```

```
mean_age = round(mean(df$age)/12)
sd_age = round(sd(df$age)/12)
```

```
prob_male = sum(with(df,sex=="male" & seropos==1))/sum(with(df,sex=="male"))
prob_female = sum(with(df,sex=="female" & seropos==1))/sum(with(df,sex=="female"))
prob_current = sum(with(df,idu=="current"))/population_size
prob_former = sum(with(df,idu=="former"))/population_size
prob_none = sum(with(df,idu=="none"))/population_size
```

```
data_summary <- data.frame(
  Calculation = c("mean_age", "sd_age", "prob_male", "prob_female", "prob_current", "prob_former", "prob_none"),
  Result = c(mean_age, sd_age, prob_male, prob_female, prob_current, prob_former, prob_none)
)
knitr::kable(data_summary, format = "simple")
```

Calculation	Result
mean_age	18.000
sd_age	5.000
prob_male	0.292
prob_female	0.199
prob_current	0.198
prob_former	0.300
prob_none	0.503

```
# Plot
plot_df <- data.frame(idu = character(), inf_prob = numeric()) # Create a data frame for plotting

idus <- unique(df$idu)
for (x in 1:length(idus)) {
  inf_prob <- sum(with(df, idu == idus[x] & inf == 1)) / sum(with(df, idu == idus[x]))
  plot_df <- rbind(plot_df, data.frame(idu = idus[x], inf_prob = inf_prob))
} # Calculate the infection probabilities for each IDU state

ggplot(plot_df, aes(x = idu, y = inf_prob)) +
  geom_bar(stat = "identity") +
  labs(x = "IDU State", y = "Percentage Infected",
       title = "Percentage of Infected People by IDU State") +
  theme_bw()
```

