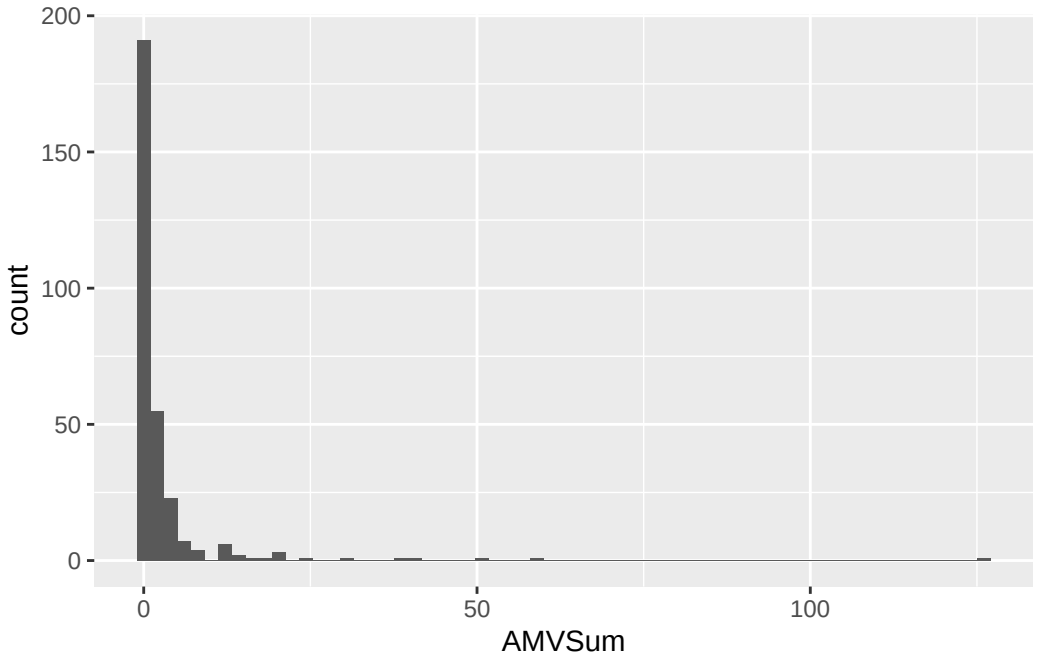
# Sim Data Analysis

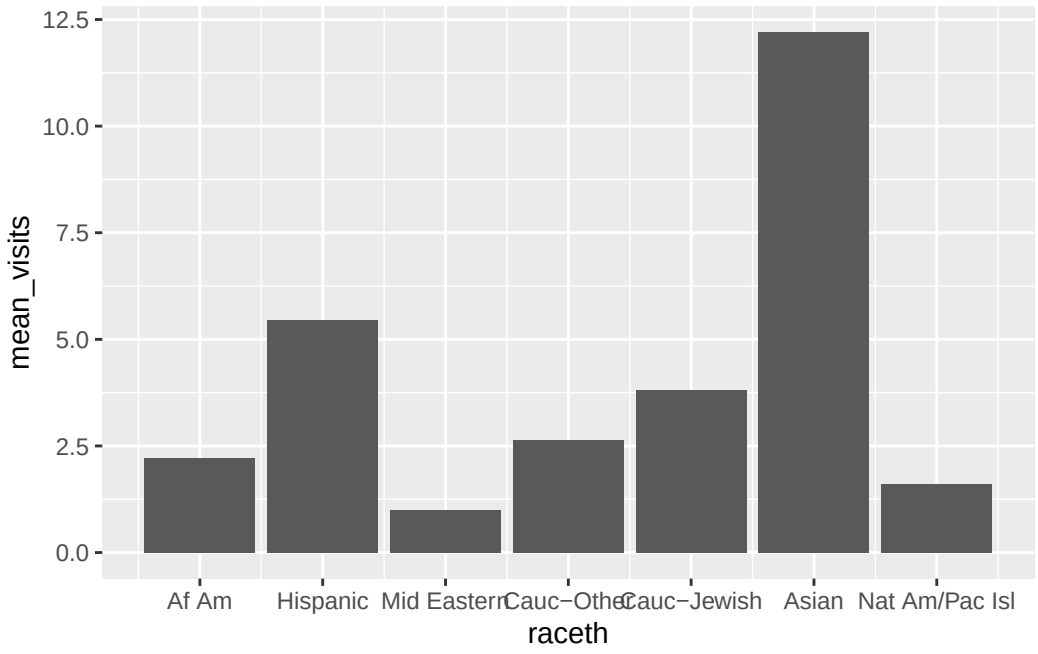
```
library(ggplot2)
library(dplyr)
```

```
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
```

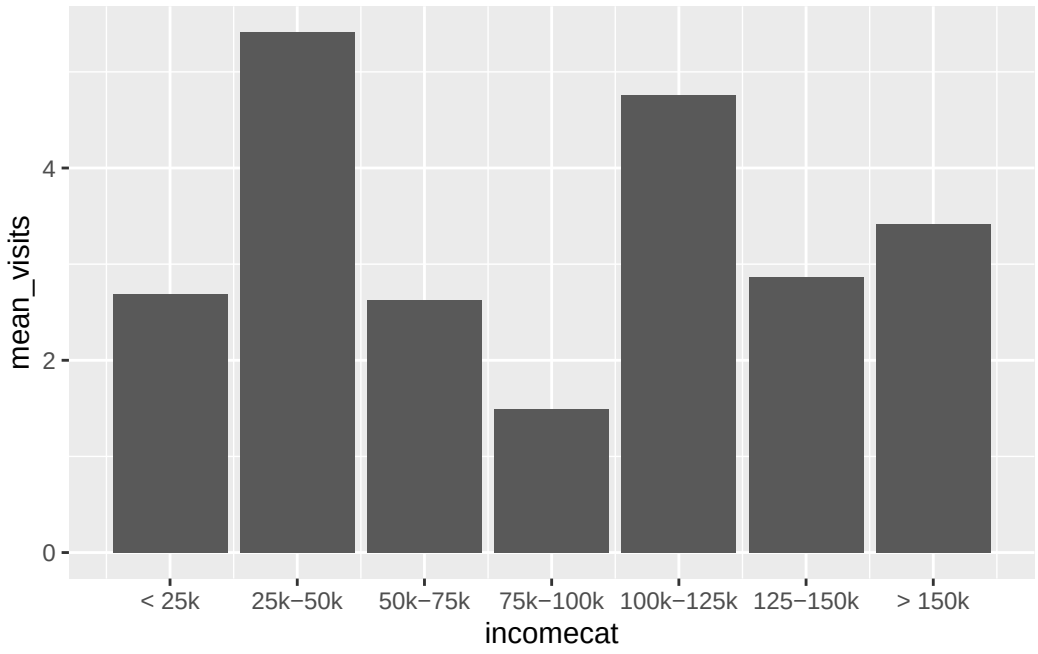
```
data <- read.csv("sim_data.csv")

# Mental health visits
data |> ggplot(aes(x = AMVSum)) +
  geom_histogram(bins=round(max(data$AMVSum)/2))
```



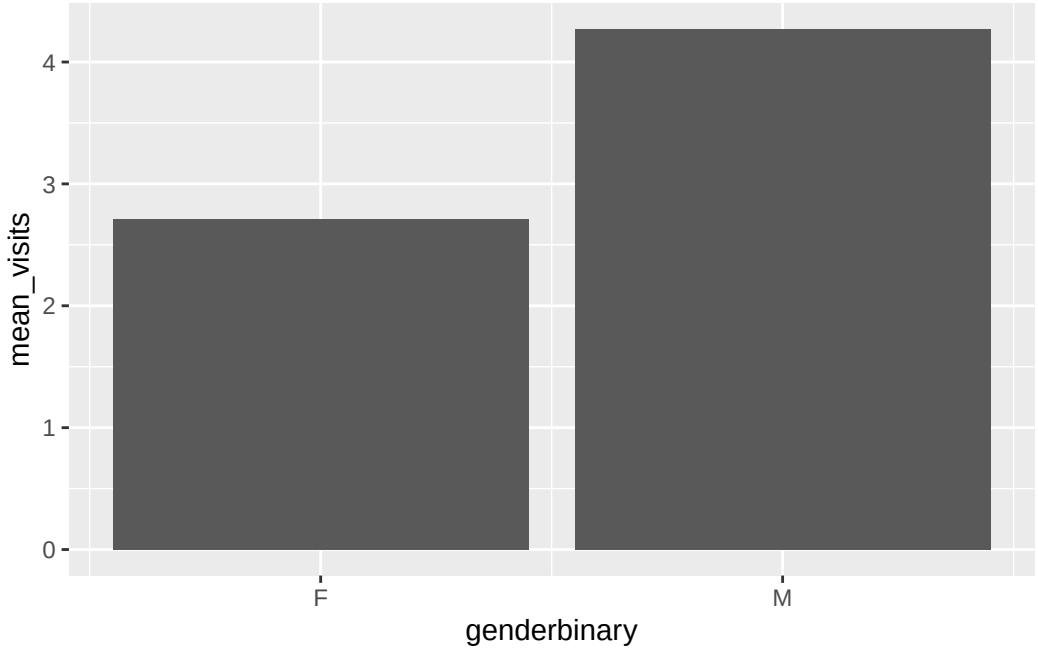

```
# # By race
data |> group_by(raceth) |>
  summarize(mean_visits = mean(AMVSum)) |>
  ggplot(aes(x = raceth, y = mean_visits)) +
  geom_bar(stat = "identity") +
  scale_x_continuous(breaks = c(1,2,3,4,5,6,7), labels = c("Af Am", "Hispanic", "Mid Easte
```



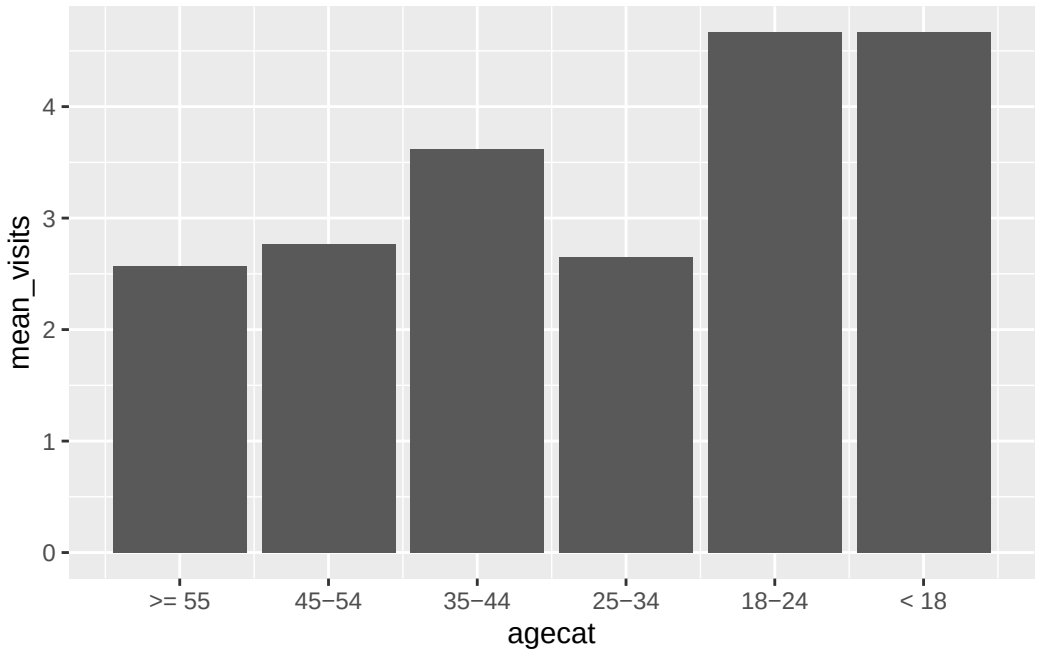

```
# By income
data |> group_by(incomecat) |>
  summarize(mean_visits = mean(AMVSum)) |>
  ggplot(aes(x = incomecat, y = mean_visits)) +
  geom_bar(stat = "identity") +
  scale_x_continuous(breaks = c(1,2,3,4,5,6,7), labels = c("< 25k", "25k-50k", "50k-75k",
```




```
# By gender
data |> group_by(genderbinary) |>
  summarize(mean_visits = mean(AMVSum)) |>
  ggplot(aes(x = genderbinary, y = mean_visits)) +
  geom_bar(stat = "identity") +
  scale_x_continuous(breaks = c(0,1), labels = c("F", "M"))
```




```
# By age
data |> group_by(agecat) |>
  summarize(mean_visits = mean(AMVSum)) |>
  ggplot(aes(x = agecat, y = mean_visits)) +
  geom_bar(stat = "identity") +
  scale_x_continuous(breaks = c(1,2,3,4,5,6), labels = c(">= 55", "45-54", "35-44", "25-34
```




```r
# Model example

model1 <- lm(glm(AMVSum ~ factor(genderbinary), family="poisson", data=data))
summary(model1)
```

```
Call:
lm(formula = glm(AMVSum ~ factor(genderbinary), family = "poisson",
    data = data))

Residuals:
    Min      1Q  Median      3Q     Max
 -4.269  -2.708  -1.708  -0.708 121.731

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)            2.7083     0.6972   3.885 0.000126 ***
factor(genderbinary)1  1.5602     1.1619   1.343 0.180376
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.66 on 298 degrees of freedom
```

```
Multiple R-squared:  0.006014,  Adjusted R-squared:  0.002678
F-statistic: 1.803 on 1 and 298 DF,  p-value: 0.1804
```

```r
exp(model1$coefficients)
```

```
         (Intercept) factor(genderbinary)1
           15.004248              4.759703
```

Exponentiated gender coefficient (incident rate ratio) should be about $1/1.1 = 0.91$.