Sim Data Analysis

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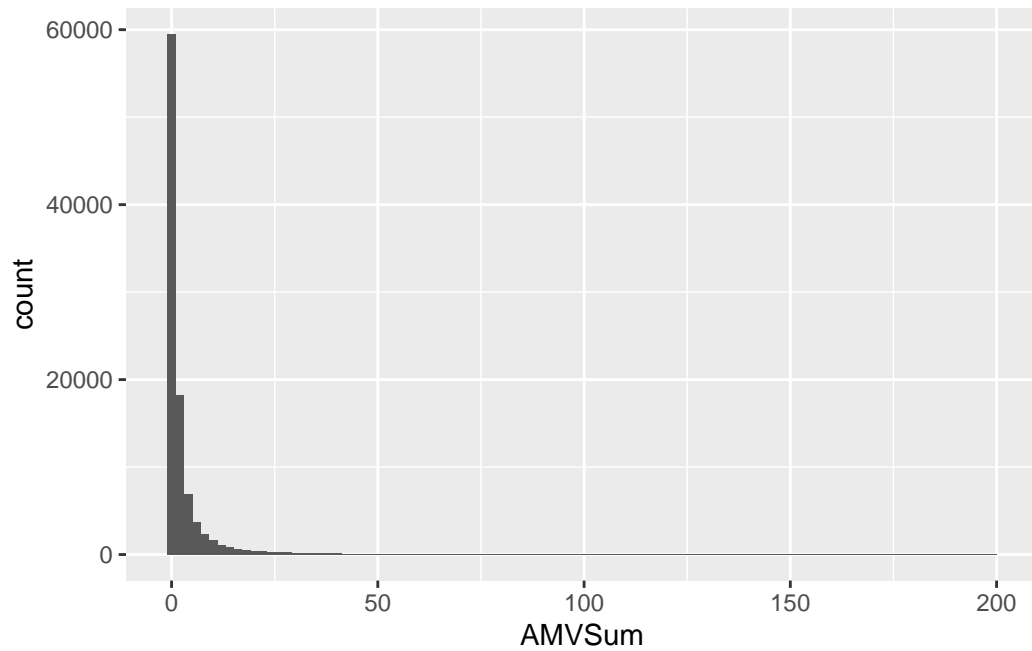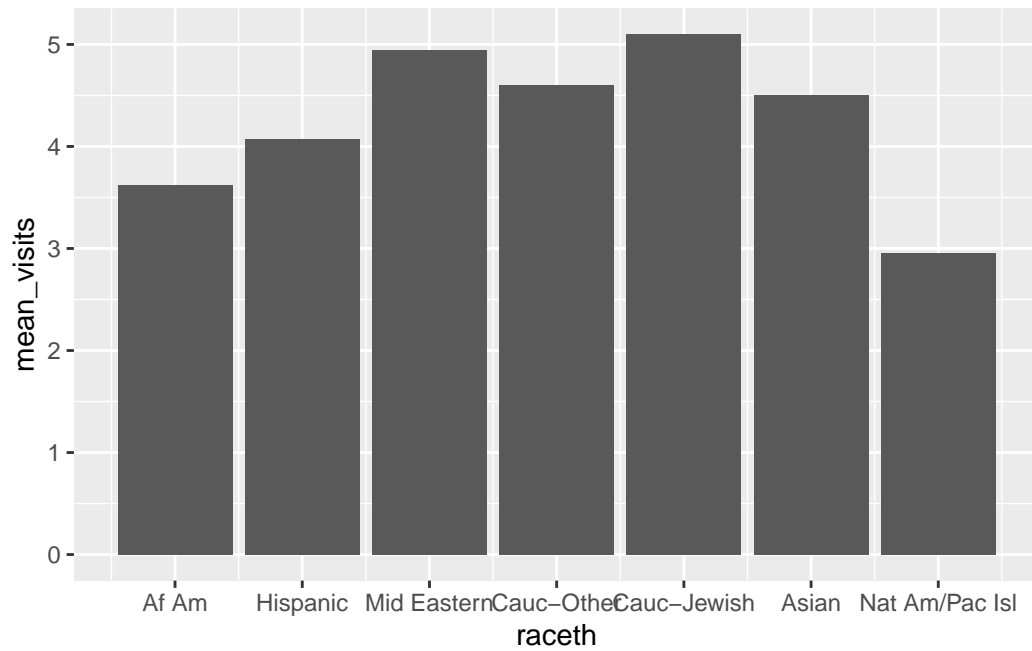
```r
data <- read.csv("sim_data.csv")

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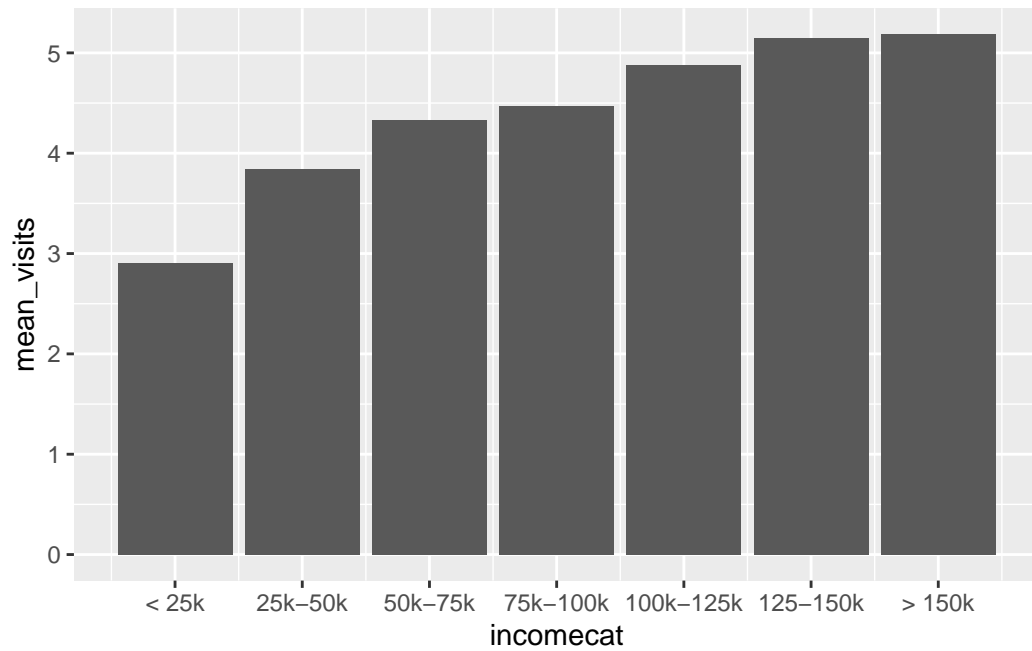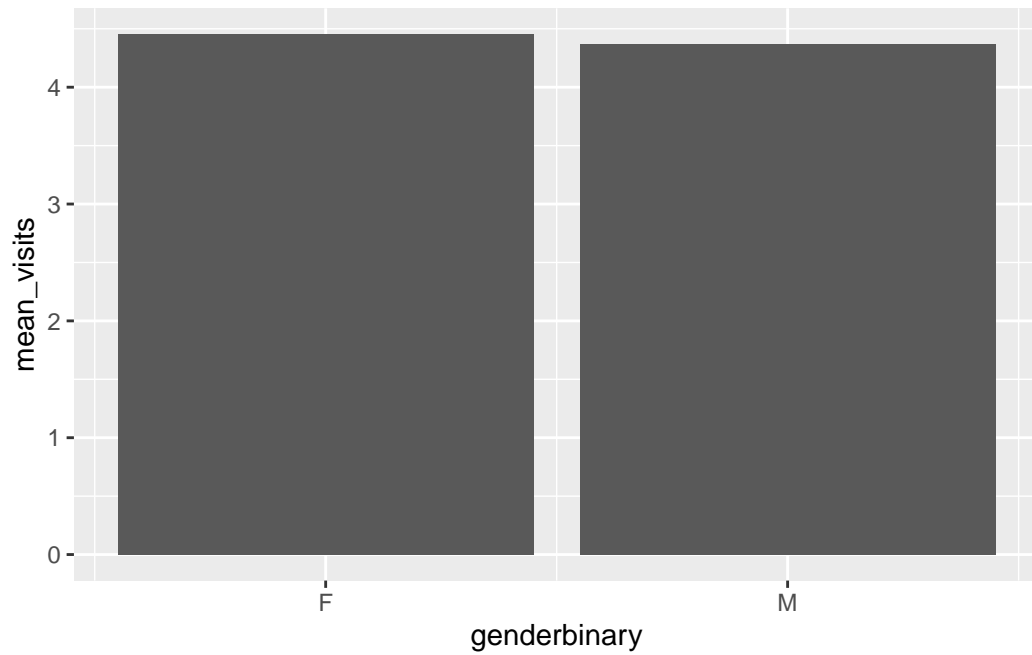
```
# # By race
data |> group_by(raceth) |>
  summarize(mean_visits = mean(AMVSum)) |>
  ggplot(aes(x = raceth, y = mean_visits)) +
  geom_bar(stat = "identity") +
  scale_x_continuous(breaks = c(1,2,3,4,5,6,7), labels = c("Af Am", "Hispanic", "Mid Easte
```

```
# 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",
```

```r
# 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
```
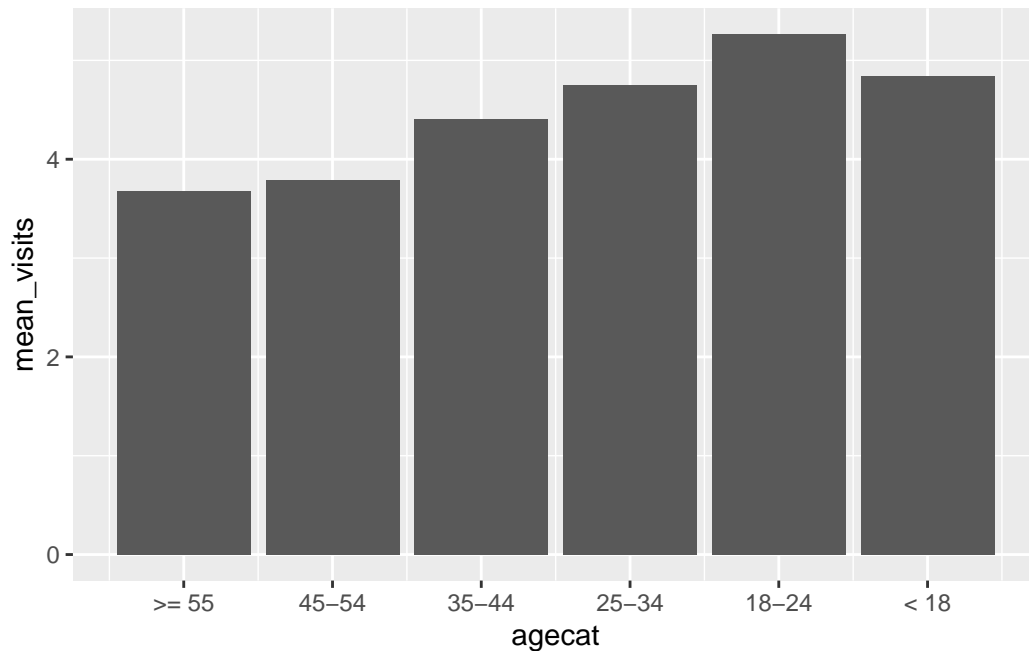
```
# Model example

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
 -4.453  -4.369  -3.369  -1.453 194.631

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)             4.45297    0.05383  82.730   <2e-16 ***
factor(genderbinary)1  -0.08381    0.08618  -0.972    0.331
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.29 on 99998 degrees of freedom
```

```
Multiple R-squared:  9.458e-06, Adjusted R-squared:  -5.425e-07
F-statistic: 0.9457 on 1 and 99998 DF,  p-value: 0.3308
```

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

```
        (Intercept) factor(genderbinary)1
         85.8820081             0.9196097
```

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