

# STAT 479 Portfolio 4

Dayne Lassiter

## MEPS

```
rm(list=ls())
require(tidyverse)
require(scales)
require(tidygraph)
require(ggraph)
require(igraph)
```

```
MEPS1 <- read_csv(file = "https://instruction.bus.wisc.edu/jffrees/jffreesbooks/Regression%20Modeling/Boo
```

Improved Data Cleaning

```
#Clean up variable names and add total expenditures variable.
```

```
MEPS <- within(MEPS1,{
  GENDER <- factor(if_else(GENDER == 0,"Male","Female"),
    levels = c("Male","Female"), ordered=TRUE)
  EXPENDTOT <- EXPENDIP + EXPENDOP
  MARISTAT <- if_else(MARISTAT1 == 0,"Never Married",
    if_else(MARISTAT1 == 1, "Married",
    if_else(MARISTAT1 == 2, "Widowed",
      "Separated"))))
  RACE <- if_else(RACE1 == 0,"Other",
    if_else(RACE1 == 1, "Asian",
    if_else(RACE1 == 2, "Black",
      if_else(RACE1 == 3, "Native",
        "White"))))
})
```

```
#Subset policyholders with positive expenditures with
#selected variables of interest.
```

```
MEPS <- MEPS[MEPS$EXPENDTOT > 0,
  c("AGE","GENDER","EXPENDIP","EXPENDOP","EXPENDTOT","RACE",
    "REGION","EDUC","MARISTAT","INCOME")]
```

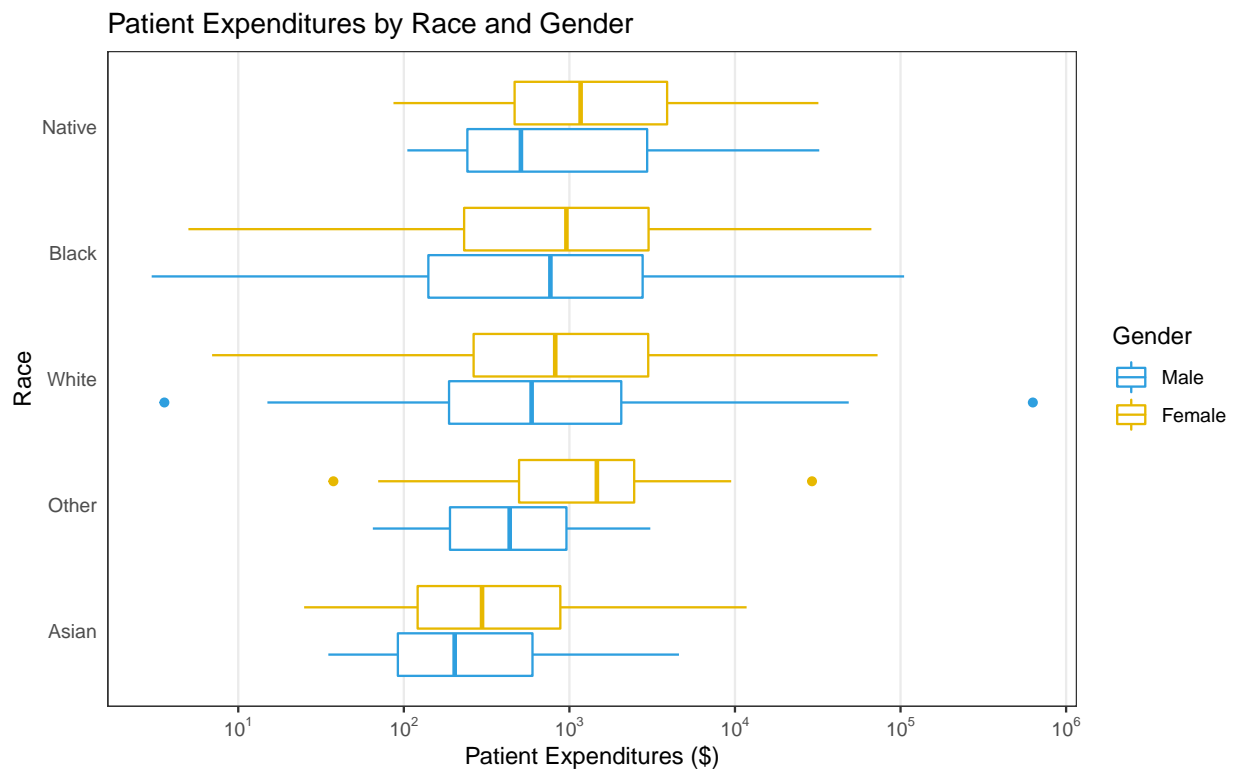
## Plot 1

```
ggplot(MEPS,aes(reorder(RACE,EXPENDTOT,IQR), EXPENDTOT, color = GENDER)) +
  geom_boxplot() +
  coord_flip() +
```

```

scale_y_log10(breaks = trans_breaks("log10", function(x) 10^x),
              labels = trans_format("log10", math_format(10^.x))) +
scale_color_manual(values = c("#2E9FDF", "#E7B800")) +
labs(
  x="Race",
  y="Patient Expenditures ($)",
  title = "Patient Expenditures by Race and Gender",
  color = "Gender"
) +
theme_bw() +
theme(
  axis.ticks.y = element_blank(),
  panel.grid.major.y = element_blank(),
  panel.grid.minor.x = element_blank()
)

```



## Plot 2

```

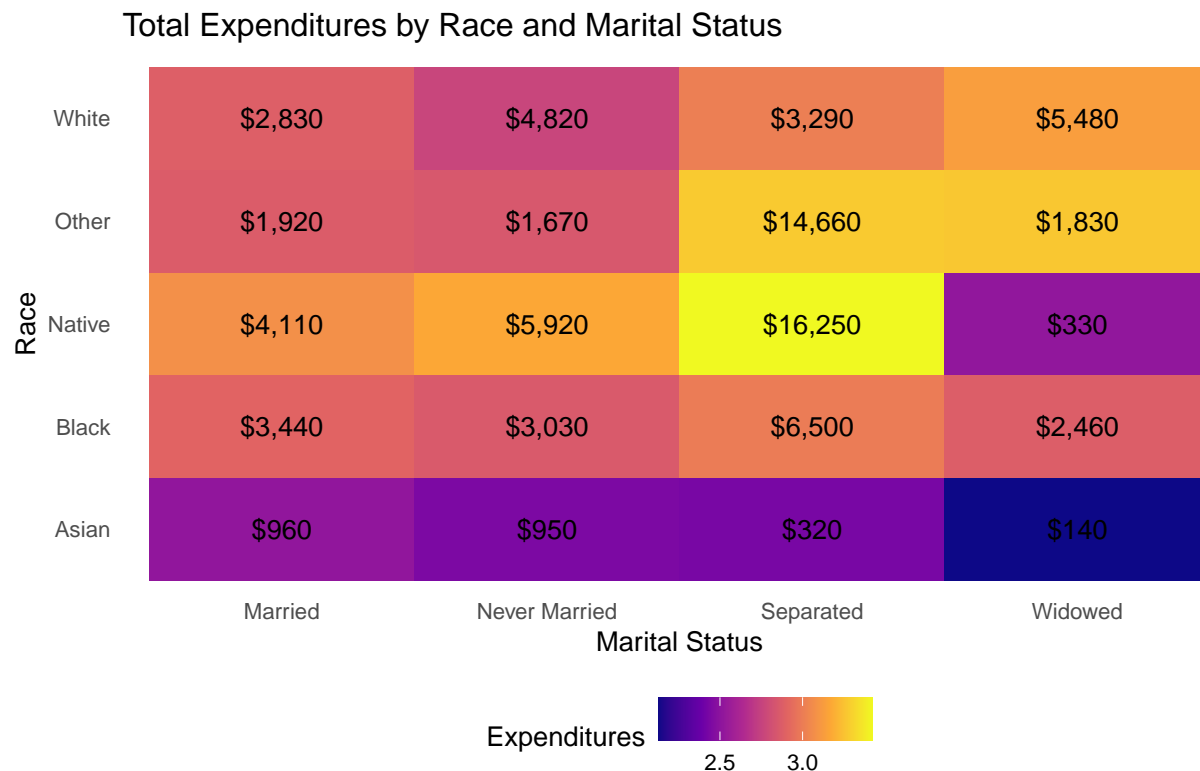
#Summarize by Race, Marital Status, and
#mean of expenditures and log(expenditures)
MEPS %>%
  group_by(RACE, MARISTAT) %>%
  #Converted the text label to dollar format with rounding
  summarize(Mean = dollar(round(mean(EXPENDTOT), -1)),
             lMean = mean(log10(EXPENDTOT))) %>%
  #Tile Plot:

```

```

ggplot(aes(MARISTAT,RACE, fill = lMean)) +
  geom_tile(aes(height = 1, width = 1)) +
  geom_text(aes(label=Mean)) +
  scale_fill_viridis_c(option = "plasma") +
  labs(
    x="Marital Status",
    y="Race",
    title = "Total Expenditures by Race and Marital Status",
    fill = "Expenditures"
  ) +
  theme(
    panel.background = element_blank(),
    axis.ticks = element_blank(),
    legend.position = "bottom"
  )

```



**A link to the data dictionary:**

<https://instruction.bus.wisc.edu/jfrees/jfreesbooks/Regression%20Modeling/BookWebDec2010/DataDescriptions.pdf>