final_project

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

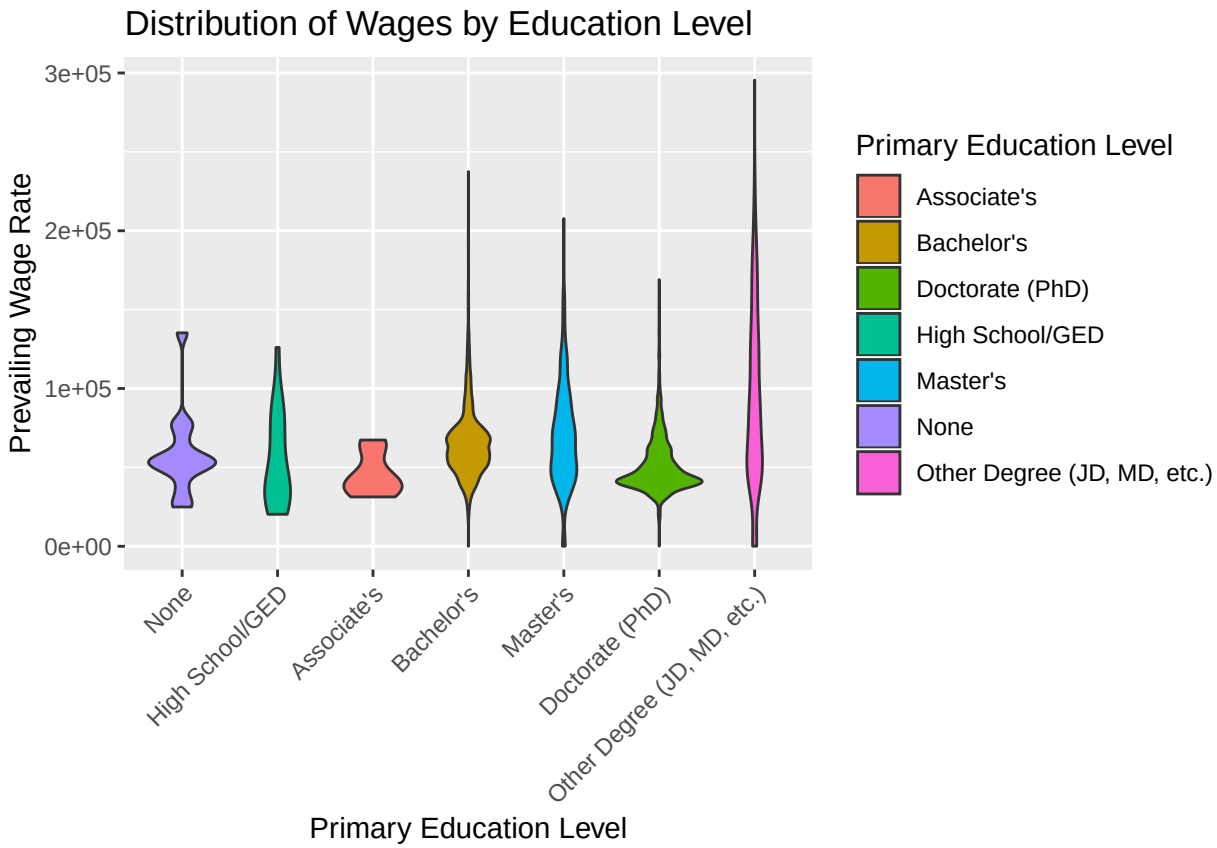
```
setwd("/Users/kevinchen/Documents/stor320")
data <- read.csv("h1data.csv", as.is=T)
```

```
wage_data <- data[!is.na(data$PWD_WAGE_RATE),]
ddply(wage_data,~PRIMARY_EDUCATION_LEVEL,summarise,mean=mean(PWD_WAGE_RATE),sd=sd(PWD_WAGE_RATE))
```

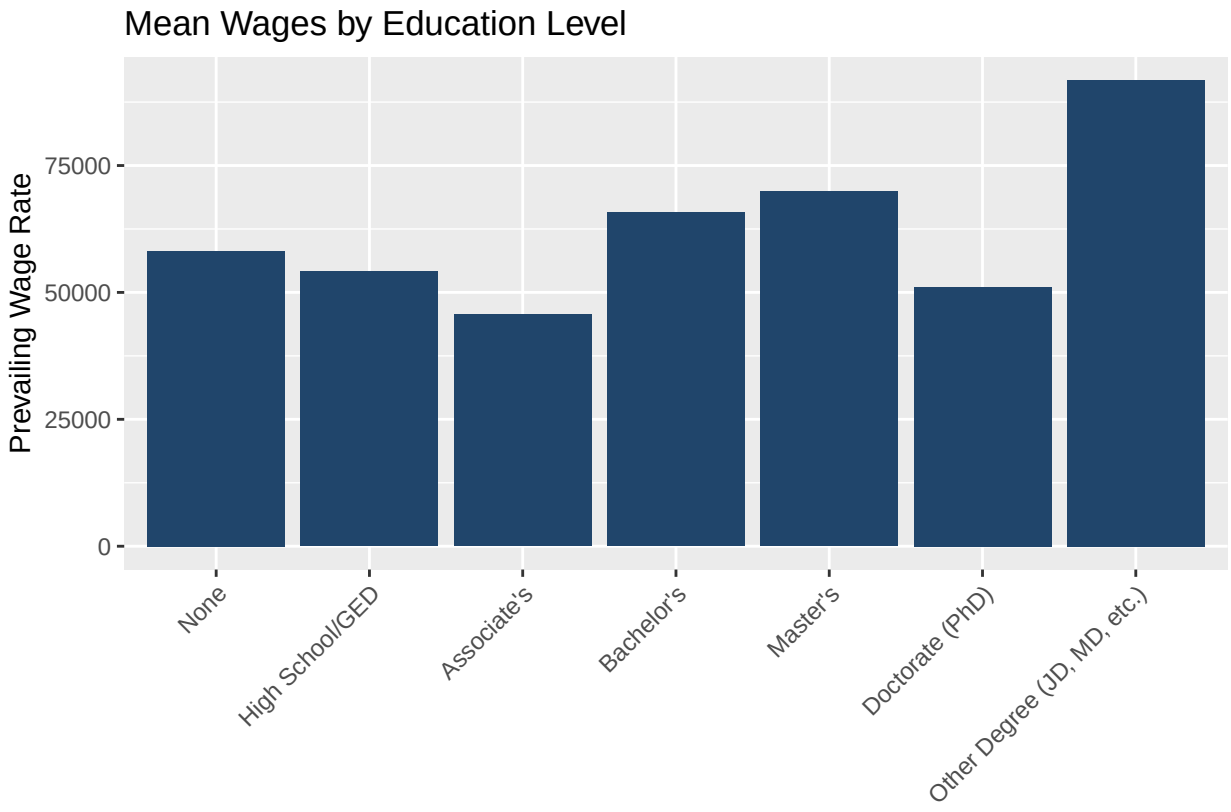
```
##       PRIMARY_EDUCATION_LEVEL     mean       sd
## 1                Associate's 45632.14 13751.67
## 2                 Bachelor's 65755.86 23033.99
## 3            Doctorate (PhD) 51056.80 17895.11
## 4            High School/GED 54175.28 29823.00
## 5                   Master's 69861.80 32404.65
## 6                       None 58133.33 26082.73
## 7 Other Degree (JD, MD, etc.) 91824.45 55562.52
```

```
level <- c("None", "High School/GED", "Associate's", "Bachelor's", "Master's", "Doctorate (PhD)", "Othe

ggplot(wage_data, aes(factor(PRIMARY_EDUCATION_LEVEL, levels = level), PWD_WAGE_RATE)) +
  geom_violin(aes(fill = factor(PRIMARY_EDUCATION_LEVEL))) +
  ggtitle("Distribution of Wages by Education Level") +
  xlab("Primary Education Level") +
  ylab("Prevailing Wage Rate") +
  labs(fill = "Primary Education Level") +
  theme(axis.text.x=element_text(angle=45, hjust=1))
```




```
ggplot(wage_data, aes(factor(PRIMARY_EDUCATION_LEVEL, levels = level), PWD_WAGE_RATE)) +
  stat_summary(fun.y="mean", geom="bar", fill = "#20456b") +
  ggtitle("Mean Wages by Education Level") +
  xlab("Primary Education Level") +
  ylab("Prevailing Wage Rate") +
  theme(axis.text.x=element_text(angle=45, hjust=1))
```




```
count <- aggregate(cbind(count = VISA_CLASS) ~ PRIMARY_WORKSITE_STATE,
          data = data,
          FUN = function(x){NROW(x)})
count <- count[-c(8,35,40,48),]
count
```

```
##    PRIMARY_WORKSITE_STATE count
## 1                 ALABAMA    52
## 2                 ARIZONA    61
## 3                ARKANSAS    19
## 4              CALIFORNIA   556
## 5                COLORADO    29
## 6             CONNECTICUT    54
## 7                DELAWARE    36
## 9                 FLORIDA   541
## 10                GEORGIA   181
## 11                 HAWAII    11
## 12                  IDAHO     2
## 13               ILLINOIS   471
## 14                INDIANA   214
## 15                   IOWA    18
## 16                 KANSAS    75
## 17               KENTUCKY    10
## 18              LOUISIANA    37
## 19                  MAINE     1
## 20               MARYLAND   119
## 21          MASSACHUSETTS   461
```

```
## 22               MICHIGAN   175
## 23              MINNESOTA    38
## 24            MISSISSIPPI    72
## 25               MISSOURI    37
## 26                MONTANA     2
## 27               NEBRASKA     6
## 28                 NEVADA     9
## 29          NEW HAMPSHIRE     4
## 30             NEW JERSEY   178
## 31             NEW MEXICO    10
## 32               NEW YORK   459
## 33         NORTH CAROLINA   155
## 34           NORTH DAKOTA    11
## 36                   OHIO   177
## 37               OKLAHOMA    11
## 38                 OREGON    22
## 39           PENNSYLVANIA   121
## 41           RHODE ISLAND    13
## 42         SOUTH CAROLINA   115
## 43           SOUTH DAKOTA    21
## 44              TENNESSEE    25
## 45                  TEXAS   964
## 46                   UTAH     9
## 47                VERMONT     2
## 49               VIRGINIA    95
## 50             WASHINGTON   161
## 51          WEST VIRGINIA    59
## 52              WISCONSIN    40
## 53                WYOMING     2
```

```
count$PRIMARY_WORKSITE_STATE <- tolower(count$PRIMARY_WORKSITE_STATE)

ggplot(count, aes(map_id = PRIMARY_WORKSITE_STATE)) +
  # map points to the fifty_states shape data
  geom_map(aes(fill = count), map = fifty_states) +
  borders("state", colour = "white") +
  scale_fill_gradient(low = "#56B1F7", high = "#132B43", guide = "colorbar") +
  expand_limits(x = fifty_states$long, y = fifty_states$lat) +
  coord_map() +
  scale_x_continuous(breaks = NULL) +
  scale_y_continuous(breaks = NULL) +
  ggtitle("Number of H1B Visa Holders by State") +
  labs(x = "", y = "", fill = "Count") +
  theme(legend.position = "bottom", panel.background = element_blank())
```

Number of H1B Visa Holders by State