

Jonestown Data Analysis

Alaine Banninga

4/13/2020

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --
## v ggplot2 3.3.0      v purrr  0.3.3
## v tibble  3.0.0      v dplyr  0.8.5
## v tidyr   1.0.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

jonestown <- read.csv(file = "data/Jonestown.csv")
```

Distribution of Jonestown Residents: Age, Gender, and Status

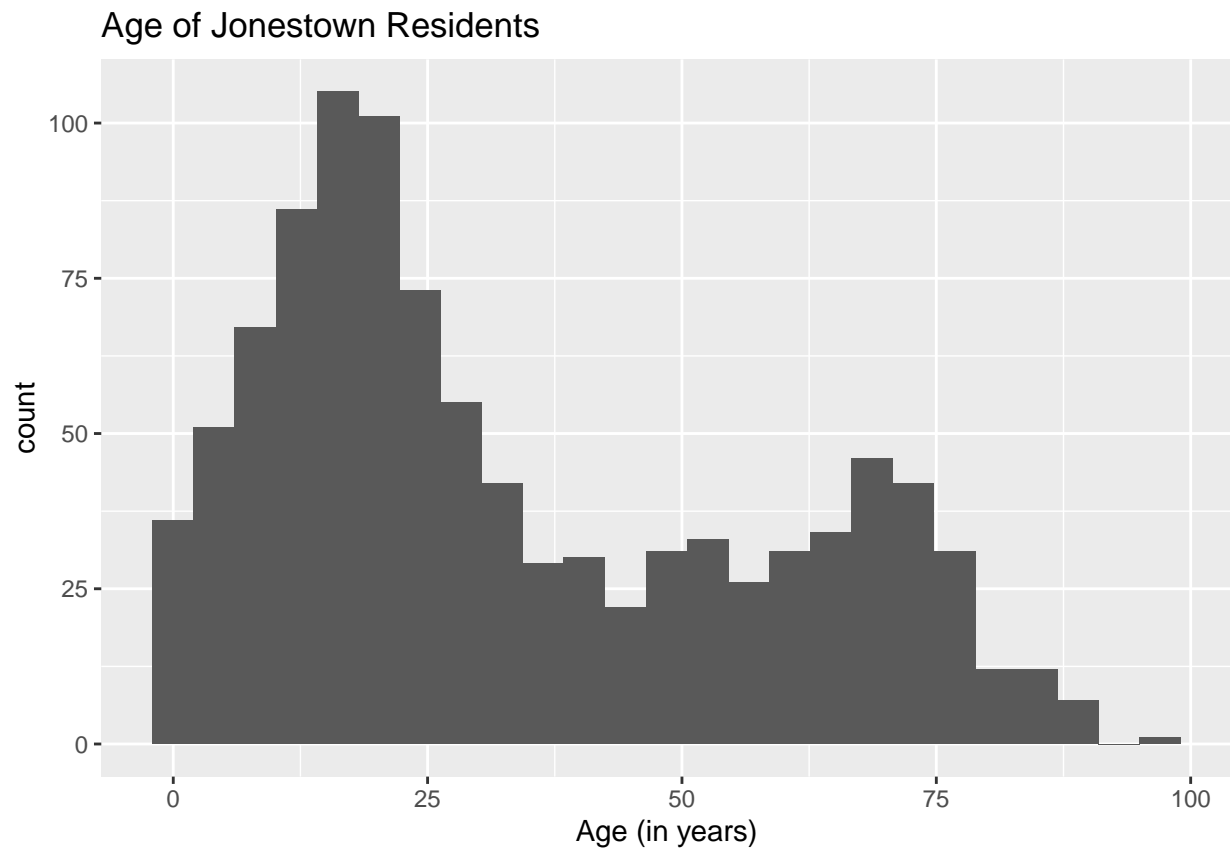
Univariate Summaries of Age, Gender, and Status

```
summary(jonestown$Age)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##      0.00   15.00   25.00   33.64   53.00   97.00         9

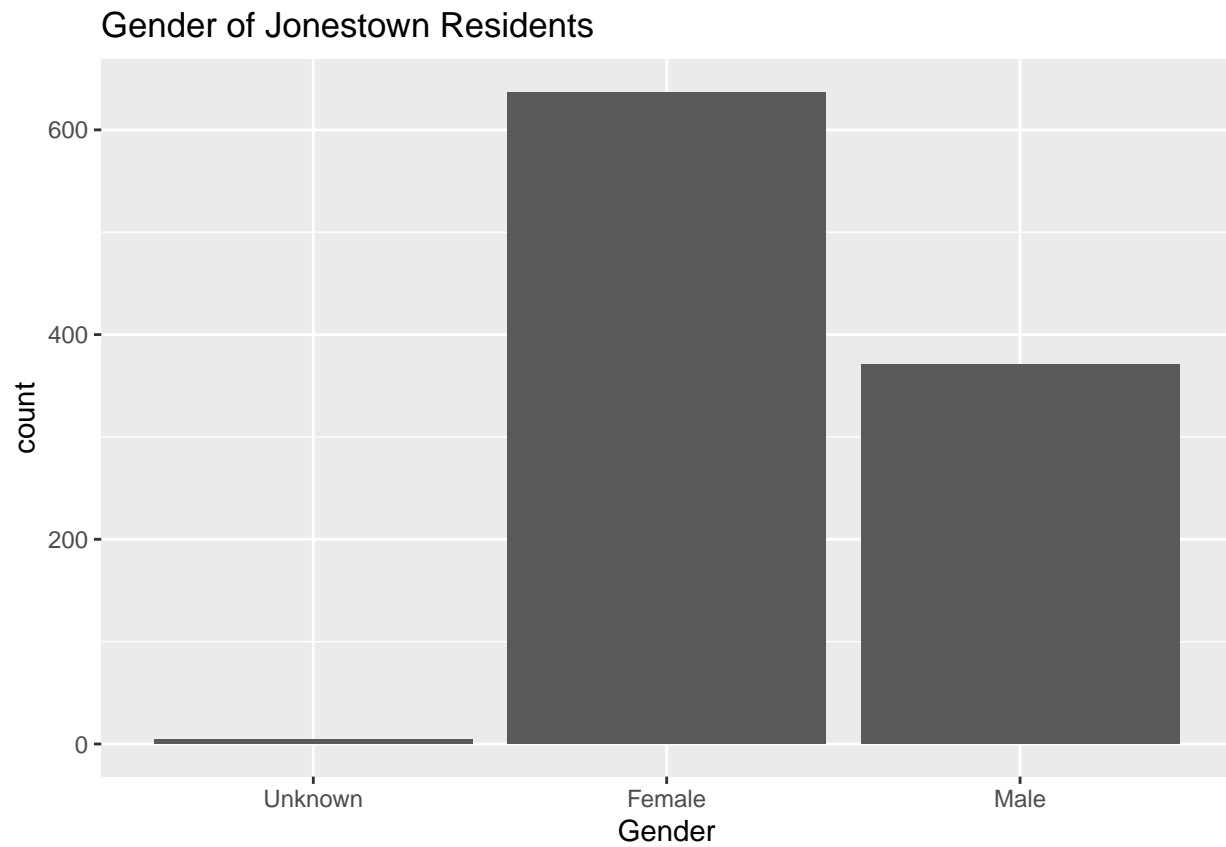
# histogram of age variable
ageHist <- ggplot(data = jonestown, aes(x = Age)) +
  geom_histogram(bins = 25) +
  ggtitle("Age of Jonestown Residents") +
  xlab("Age (in years)")
ageHist

## Warning: Removed 9 rows containing non-finite values (stat_bin).
```



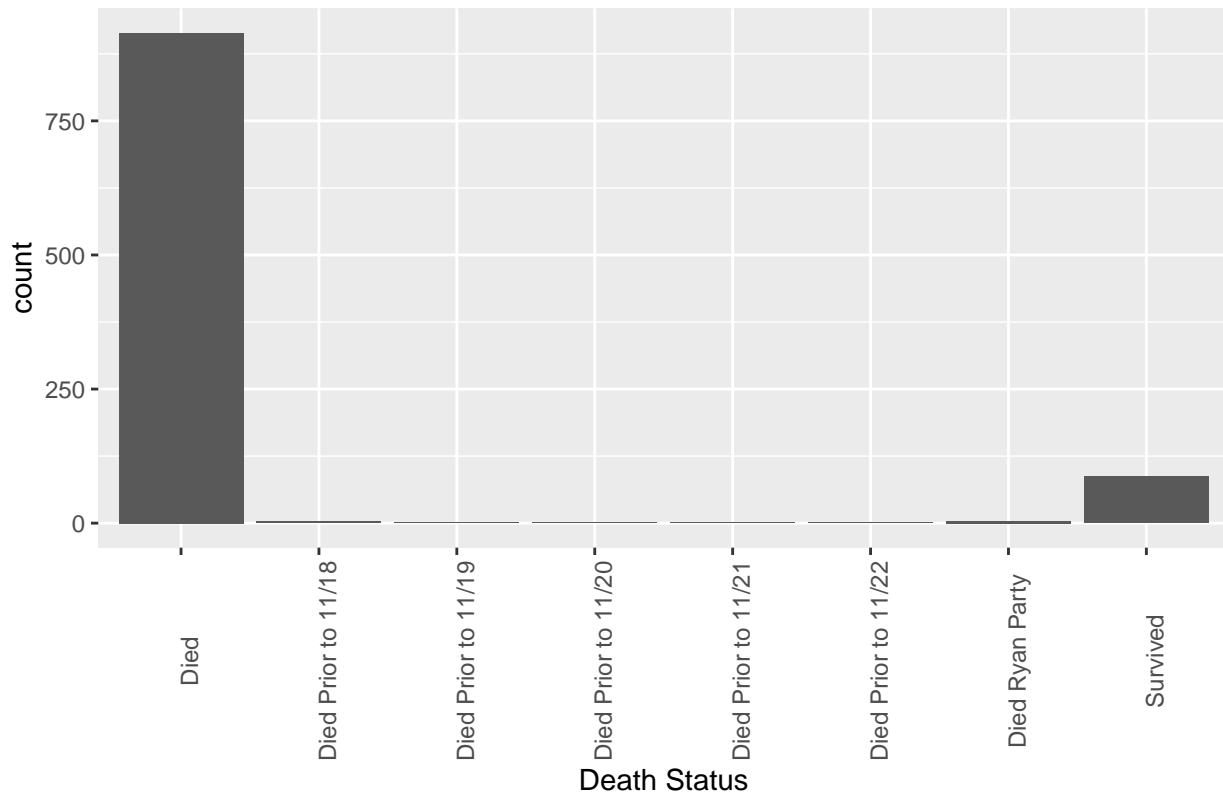
```
# relabeling gender variable
jonestown$Gender <- fct_collapse(jonestown$Gender,
  "Unknown" = c("", "u"),
  "Female" = "f",
  "Male" = "m")

# bar plot of gender
ggplot(data = jonestown, aes(x = Gender)) +
  geom_bar() +
  ggtitle("Gender of Jonestown Residents") +
  xlab("Gender")
```



```
# bar plot: death status counts (specific dates of death)
ggplot(data = jonestown, aes(x = Status)) +
  geom_bar() +
  ggtitle("Death Status of Jonestown Residents (specific dates)") +
  xlab("Death Status") +
  theme(axis.text.x = element_text(angle = 90))
```

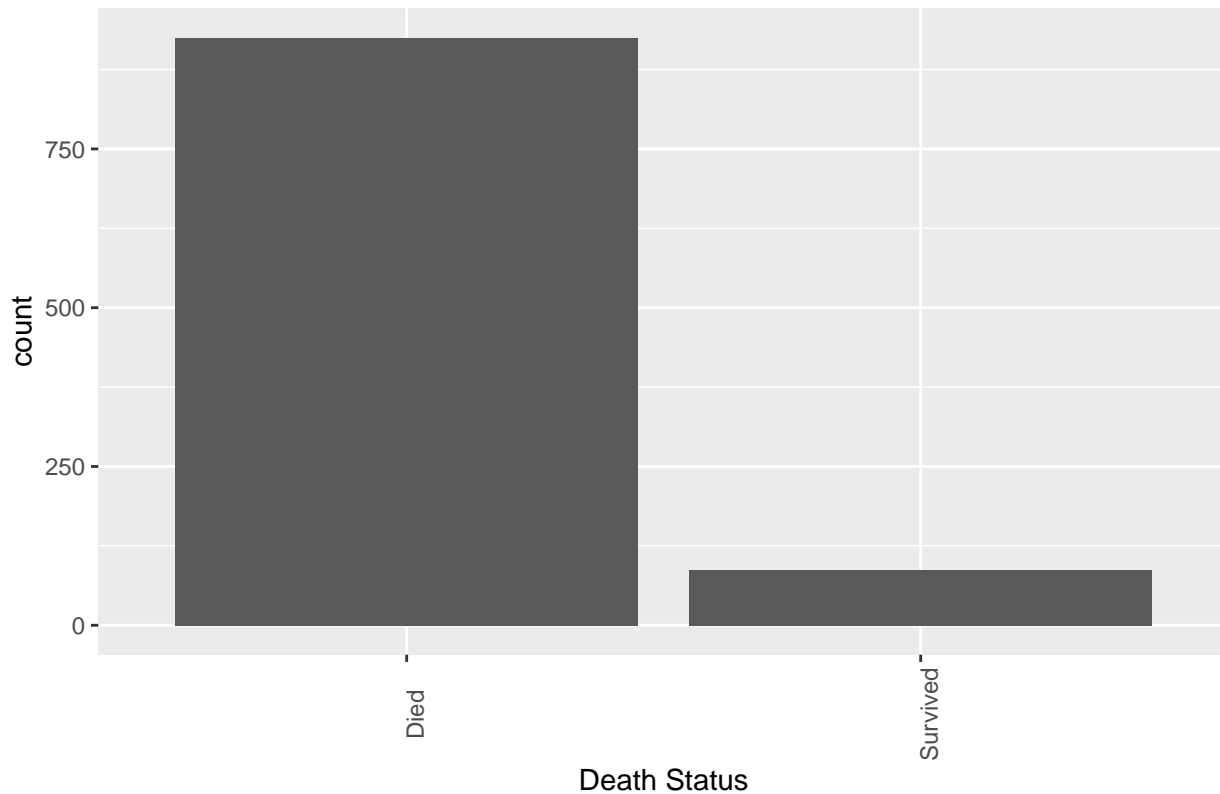
Death Status of Jonestown Residents (specific dates)



```
# Collapsing all residents listed as dead in dataset as "Died"
newStatus <- fct_collapse(jonestown$Status,
                          Died = c("Died", "Died Prior to 11/18", "Died Prior to 11/19", "Died Prior to 11/20", "Died Prior to 11/21", "Died Prior to 11/22", "Died Ryan Party"),
                          Survived = "Survived")

# bar plot: death status counts (dead or alive)
ggplot(data = jonestown, aes(x = newStatus)) +
  geom_bar() +
  ggtitle("Death Status of Jonestown Residents (Dead or Alive)") +
  xlab("Death Status") +
  theme(axis.text.x = element_text(angle = 90))
```

Death Status of Jonestown Residents (Dead or Alive)



Place of Birth by Age, Gender, and Status

```
# calculating number of residents per birth country
table(jonestown$BirthCountry)
```

```
##
##           Belize British West Indies           Canada           Germany
##              2              1              1              1
##    Great Britain           Guyana           Honduras           USA
##              1              21              1              984
```

```
mean(jonestown$Age, na.rm = TRUE) # mean age of Jonestown residents on 11/18/1978
```

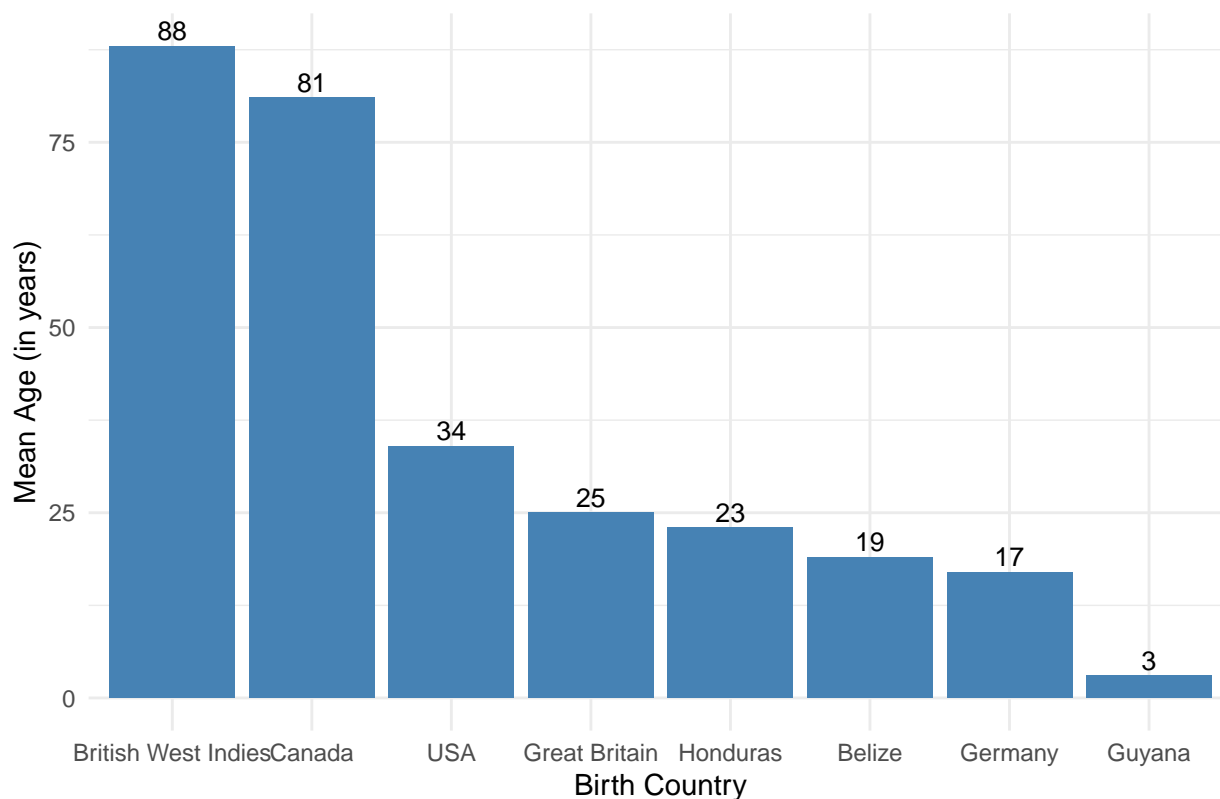
```
## [1] 33.64108
```

```
# bar plot: mean age of jonestown resident by birth country
```

```
ageByPOB <- jonestown %>%
  group_by(BirthCountry) %>%
  summarise(mean_age = round(mean(Age, na.rm = TRUE), digits = 0)) %>%
  ggplot(aes(x= reorder(BirthCountry, -mean_age), y=mean_age)) +
  geom_bar(stat = "identity", fill = "steelblue")+
  geom_text(aes(label=mean_age), vjust=-0.3, size=3.5)+
  theme_minimal() +
  ggtitle("Mean Age of Jonestown Residents by their Birth Country") +
  ylab("Mean Age (in years)") +
  xlab("Birth Country")
```

```
ageByPOB
```

Mean Age of Jonestown Residents by their Birth Country



```
t <- table(jonestown$BirthCountry, jonestown$Gender)
```

```
# Percentages of Males, Females, and Unknown Gender Residents by Country
prop.table(t,1)
```

```
##
##               Unknown      Female      Male
## Belize           0.00000000 0.50000000 0.50000000
## British West Indies 0.00000000 1.00000000 0.00000000
## Canada           0.00000000 1.00000000 0.00000000
## Germany           0.00000000 1.00000000 0.00000000
## Great Britain     0.00000000 0.00000000 1.00000000
## Guyana            0.09523810 0.42857143 0.47619048
## Honduras          0.00000000 0.00000000 1.00000000
## USA               0.00203252 0.63414634 0.36382114
```

```
t2 <- table(jonestown$BirthCountry, newStatus)
```

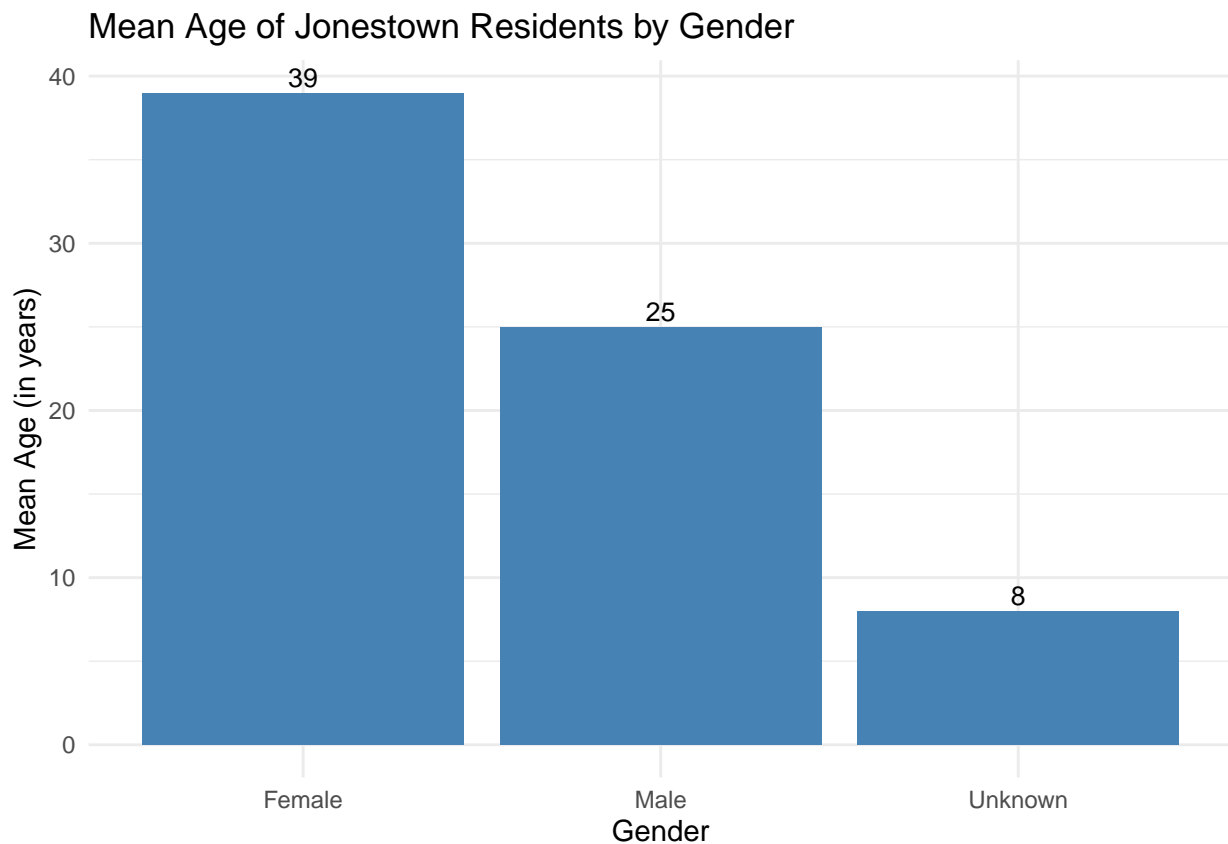
```
# Death Status Percentages by Country
prop.table(t2,1)
```

```
##               newStatus
##               Died   Survived
## Belize           1.00000000 0.00000000
## British West Indies 1.00000000 0.00000000
## Canada           1.00000000 0.00000000
## Germany           1.00000000 0.00000000
## Great Britain     0.00000000 1.00000000
```

```
## Guyana          1.00000000 0.00000000
## Honduras        1.00000000 0.00000000
## USA              0.91260163 0.08739837
```

Gender vs Age

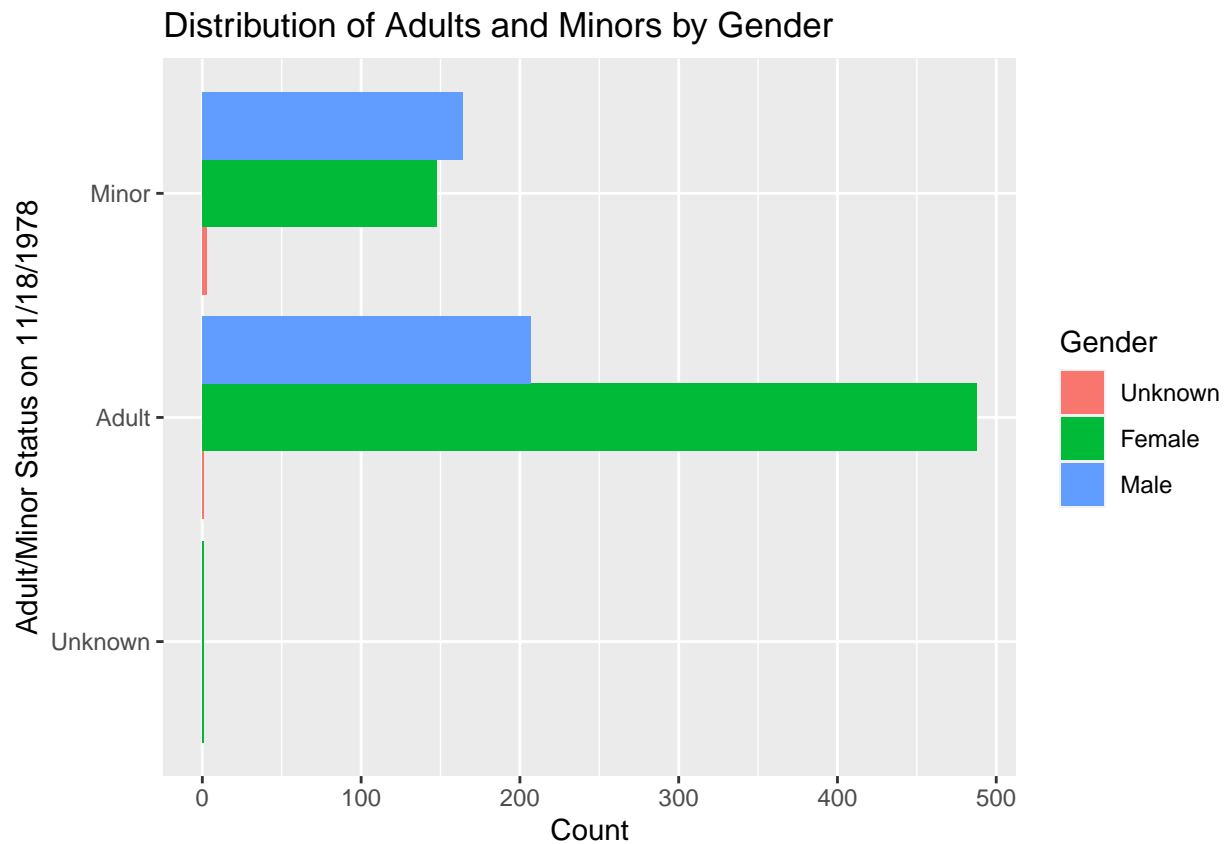
```
# bar plot: mean age of jonestown residents by gender
jonestown %>%
  group_by(Gender) %>%
  summarise(mean_age = round(mean(Age, na.rm = TRUE), digits = 0)) %>%
  ggplot(aes(x=reorder(Gender, -mean_age), y=mean_age)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  geom_text(aes(label=mean_age), vjust=-0.3, size=3.5) +
  theme_minimal() +
  ggtitle("Mean Age of Jonestown Residents by Gender") +
  ylab("Mean Age (in years)") +
  xlab("Gender")
```



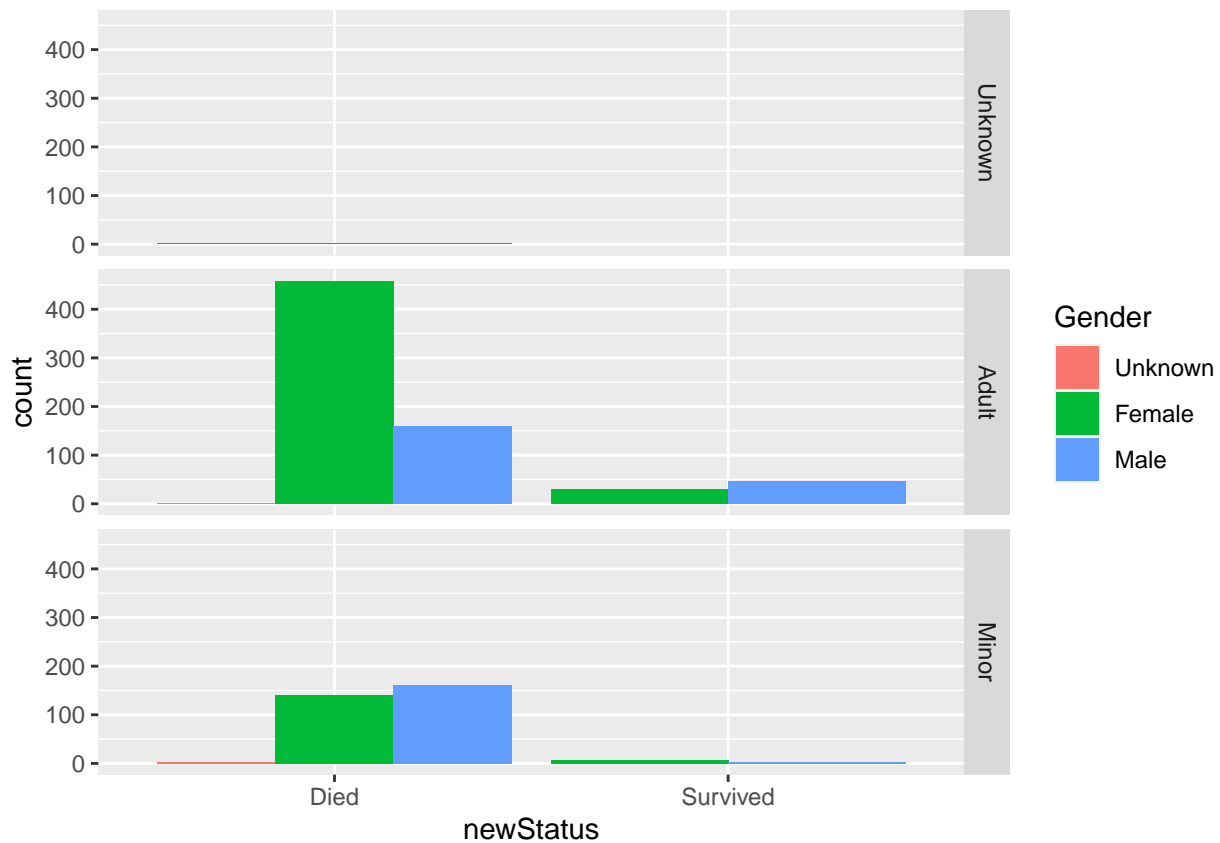
```
# label blank values as "unknown"
jonestown$AdultMinor <- fct_collapse(jonestown$AdultMinor,
  "Unknown" = "")

# bar plot showing number of adults and minors by gender
ggplot(jonestown, aes(x = AdultMinor, fill = Gender)) +
  geom_bar(position = position_dodge()) +
  coord_flip() +
```

```
ggtitle("Distribution of Adults and Minors by Gender") +
ylab("Count") +
xlab("Adult/Minor Status on 11/18/1978") +
labs(fill = "Gender")
```



```
# bar plot: count of each gender by death status, separated by adult/minor status
ggplot(data=jonestown, aes(x = newStatus, fill = Gender)) +
  geom_bar(stat = "count", position = position_dodge()) +
  facet_grid(AdultMinor ~ .)
```

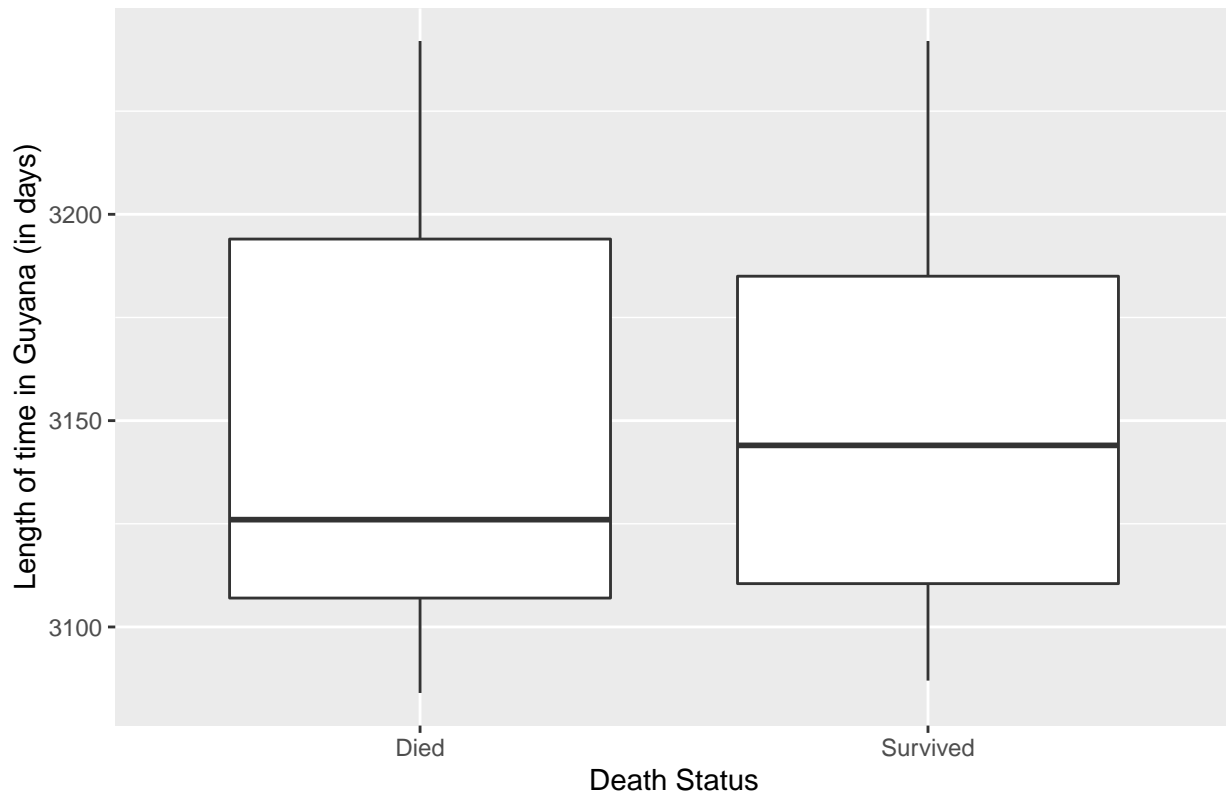
Guyana Entry vs Status

```
# num days spent in Guyana total
jonestown$daysSpent <- as.numeric(as.Date("1978-11-18") - jonestown$GuyanaEntry)

## Warning: Incompatible methods ("-.Date", "Ops.factor") for "-

# box plots of length of time spent in Guyana by death status
ggplot(data = jonestown, aes(x = newStatus, y = daysSpent)) +
  geom_boxplot() +
  ggtitle("Length of Time Spent in Guyana vs Death Status") +
  xlab("Death Status") +
  ylab("Length of time in Guyana (in days)")
```

Length of Time Spent in Guyana vs Death Status



```
f <- table(jonestown$Location, newStatus)
```

```
# Location on 11/18/1978 vs death status
prop.table(f,1)
```

```
##               newStatus
##               Died   Survived
## Albatross (Boat)  0.00000000 1.00000000
## Caracas, Venezuela 0.00000000 1.00000000
## Georgetown, Guyana 0.11904762 0.88095238
## Jonestown, Guyana  0.98704104 0.01295896
## Jonestown, Guyana* 1.00000000 0.00000000
## Port Kaituma, Guyana 0.26315789 0.73684211
```

```
# Most residents not in Jonestown, Guyana on 11/18/1978 survived
```

Jonestown Residence vs Status

```
# collapsing JTResidence into general types
# e.g. Cottage 1 and Cottage 2 are now "Cottage"
jonestown$JTResidence <- fct_collapse(jonestown$JTResidence,
                                     "Not Known" = c("", "Not Known"))
```

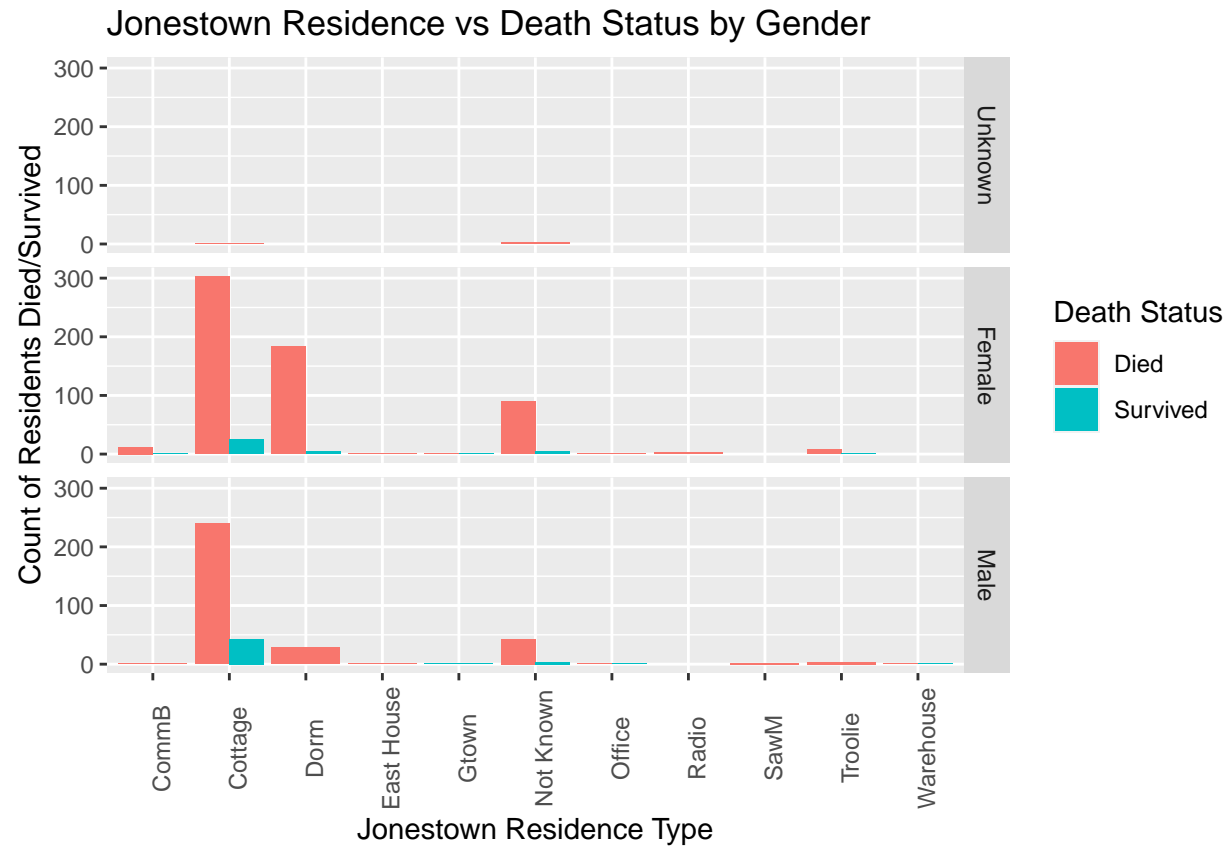
```
## Warning: Unknown levels in `f`: Not Known
```

```
jonestown$genRes <- trimws(gsub('[:digit:]]+', '', jonestown$JTResidence))
jonestown$genRes <- fct_collapse(jonestown$genRes,
                                "Cottage" = c("Cottage", "Cottage C"))

# percentage bar plot of jonestown residence type vs death status
ggplot(data = jonestown, aes(x = genRes, fill = newStatus)) +
  geom_bar(position = "fill") +
  ggtitle("Jonestown Residence vs Death Status") +
  xlab("Jonestown Residence Type") +
  theme(axis.text.x = element_text(angle = 90)) +
  ylab("Percentage of Residents Died/Survived") +
  labs(fill = "Death Status")
```



```
# bar plots of jonestown residence vs death status, separated by gender
ggplot(data=jonestown, aes(x = genRes, fill = newStatus)) +
  geom_bar(stat = "count", position = position_dodge()) +
  facet_grid(Gender ~ .) +
  theme(axis.text.x = element_text(angle = 90)) +
  ggtitle("Jonestown Residence vs Death Status by Gender") +
  xlab("Jonestown Residence Type") +
  ylab("Count of Residents Died/Survived") +
  labs(fill = "Death Status")
```

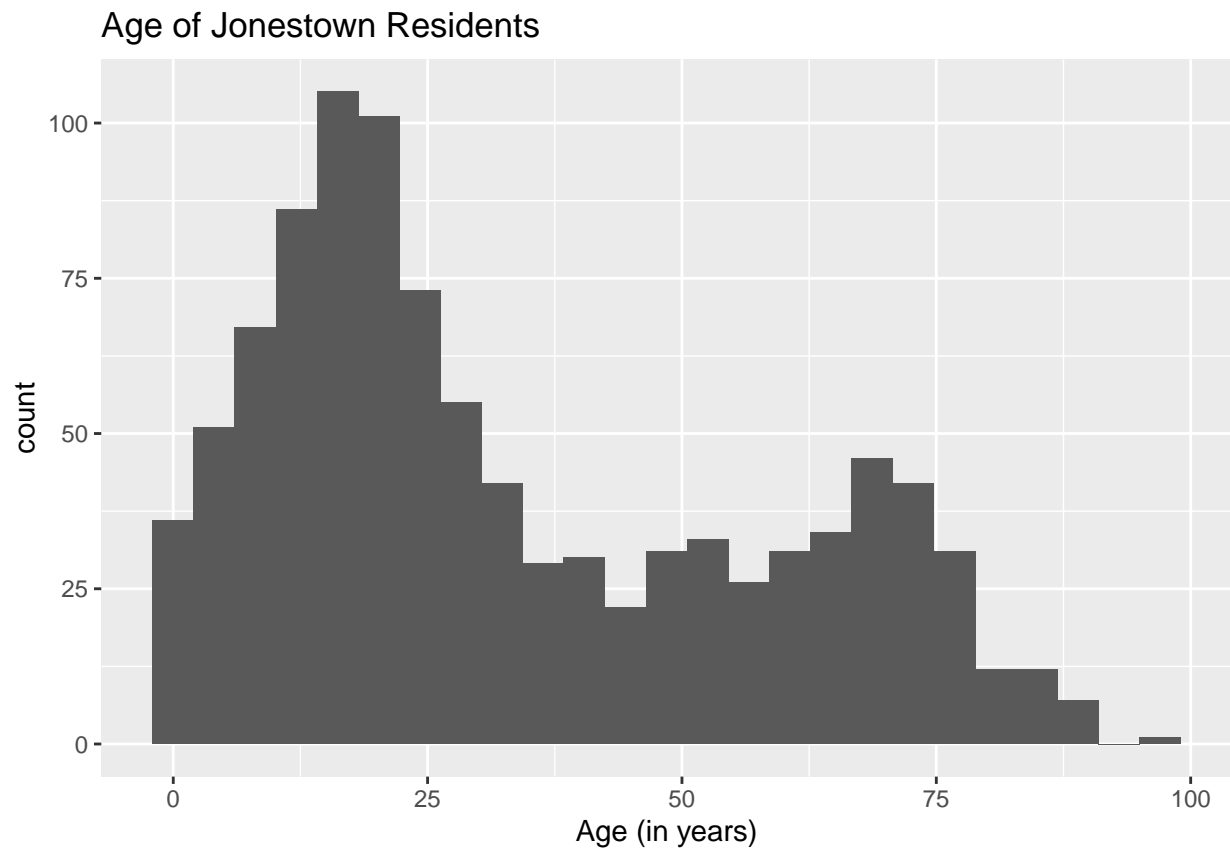


Place of Birth and its Relationship to Age

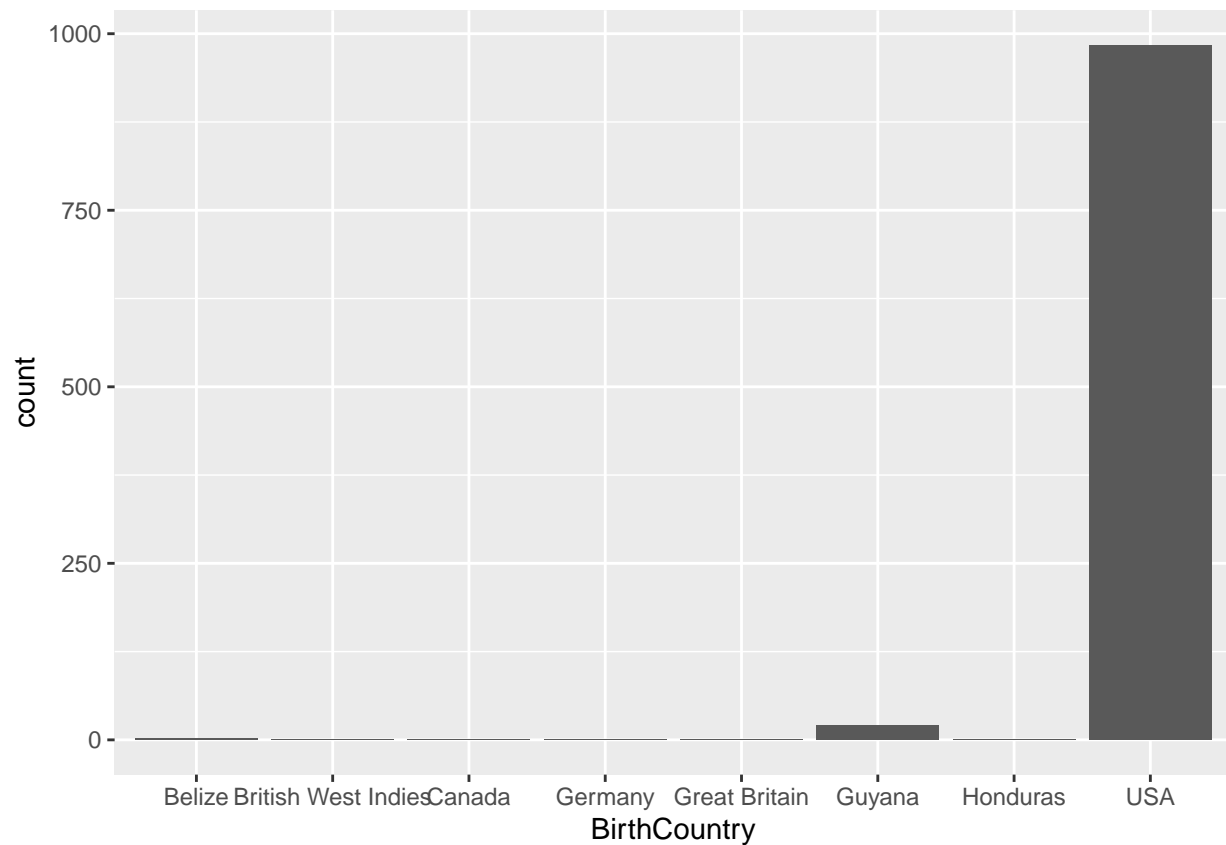
Univariate Summaries of Age, Birth Country, and Birth State

```
# histogram of age
ageHist
```

```
## Warning: Removed 9 rows containing non-finite values (stat_bin).
```



```
# not very useful, most from USA  
ggplot(data = jonestown, aes(x = BirthCountry)) +  
  geom_bar(stat = "count")
```

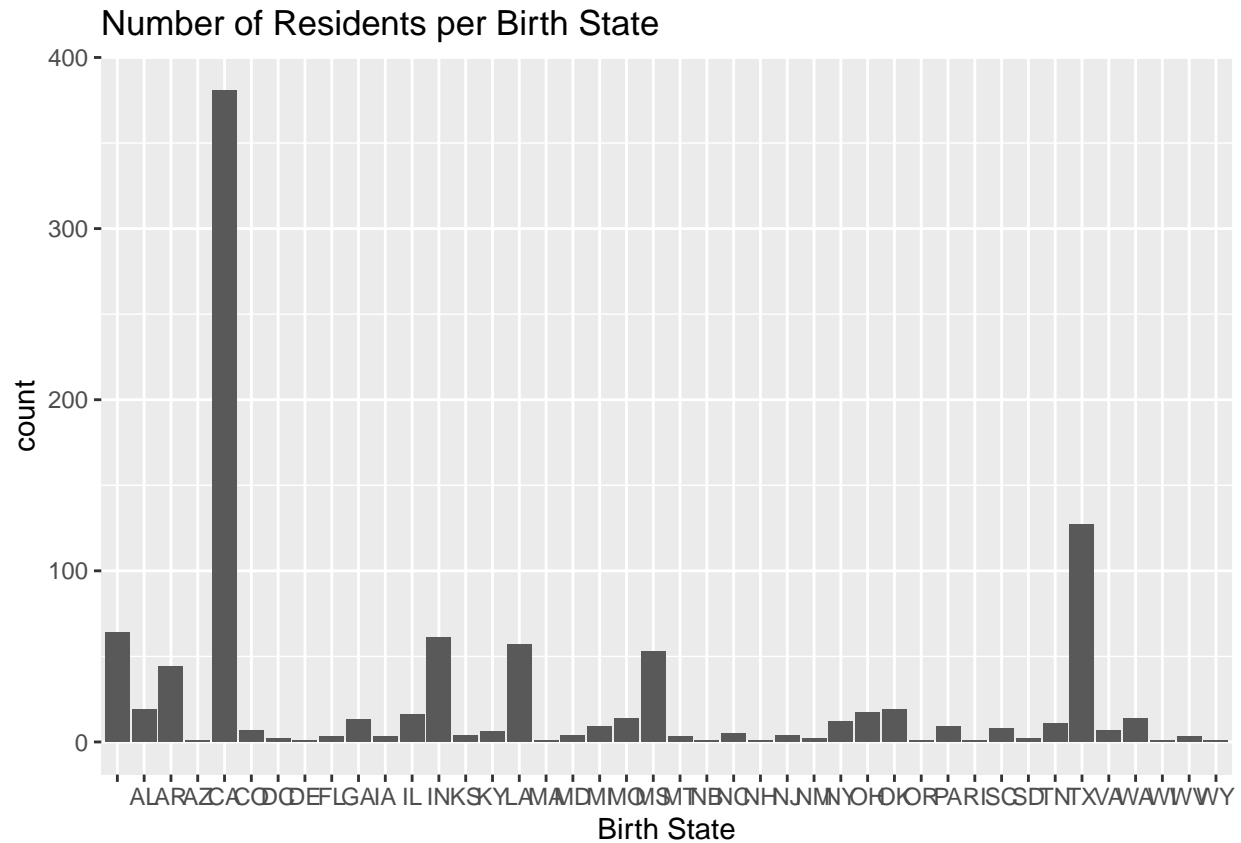


```
prop.table(table(jonestown$BirthCountry))
```

```
##
##      Belize British West Indies      Canada      Germany
##      0.0019762846      0.0009881423      0.0009881423      0.0009881423
##      Great Britain      Guyana      Honduras      USA
##      0.0009881423      0.0207509881      0.0009881423      0.9723320158
```

```
# USA overwhelming majority, next largest proportion is Guyana at 2%
```

```
ggplot(data = jonestown, aes(x = BirthState)) +
  geom_bar(stat = "count") +
  ggtitle("Number of Residents per Birth State") +
  xlab("Birth State")
```



```
prop.table(table(jonestown$BirthState))
```

```
##
##           AL           AR           AZ           CA           CO
## 0.0632411067 0.0187747036 0.0434782609 0.0009881423 0.3764822134 0.0069169960
##           DC           DE           FL           GA           IA           IL
## 0.0019762846 0.0009881423 0.0029644269 0.0128458498 0.0029644269 0.0158102767
##           IN           KS           KY           LA           MA           MD
## 0.0602766798 0.0039525692 0.0059288538 0.0563241107 0.0009881423 0.0039525692
##           MI           MO           MS           MT           NB           NC
## 0.0088932806 0.0138339921 0.0523715415 0.0029644269 0.0009881423 0.0049407115
##           NH           NJ           NM           NY           OH           OK
## 0.0009881423 0.0039525692 0.0019762846 0.0118577075 0.0167984190 0.0187747036
##           OR           PA           RI           SC           SD           TN
## 0.0009881423 0.0088932806 0.0009881423 0.0079051383 0.0019762846 0.0108695652
##           TX           VA           WA           WI           WV           WY
## 0.1254940711 0.0069169960 0.0138339921 0.0009881423 0.0029644269 0.0009881423
```

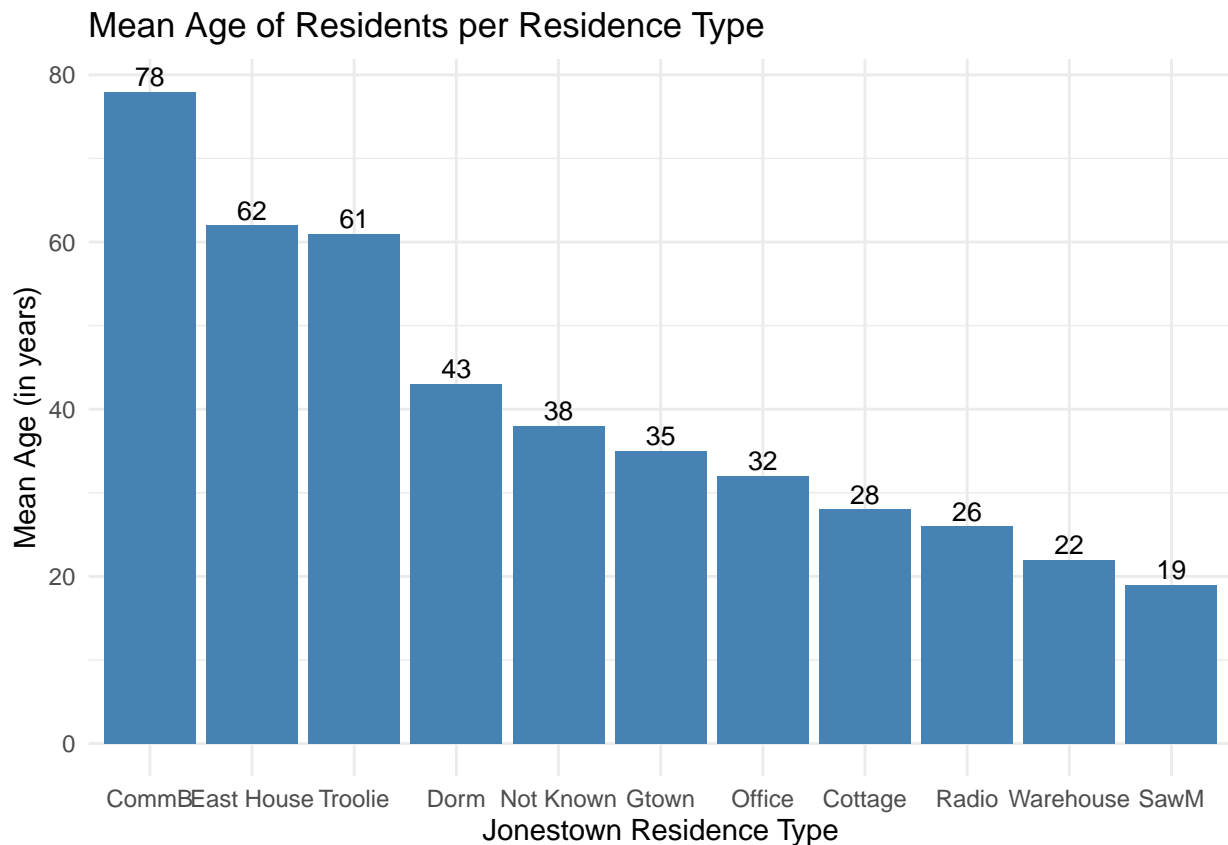
```
table(jonestown$BirthState)
```

```
##
##      AL  AR  AZ  CA  CO  DC  DE  FL  GA  IA  IL  IN  KS  KY  LA  MA  MD  MI  MO
## 64  19  44   1 381   7   2   1   3  13   3  16  61   4   6  57   1   4   9  14
## MS  MT  NB  NC  NH  NJ  NM  NY  OH  OK  OR  PA  RI  SC  SD  TN  TX  VA  WA  WI
## 53   3   1   5   1   4   2  12  17  19   1   9   1   8   2  11 127   7  14   1
## WV  WY
##   3   1
```

```
# highest proportions from CA, TX; lowest from AZ, DE, MA, NB, OR, RI, WI, WY
```

Mean Age of Residents by their Jonestown Residence

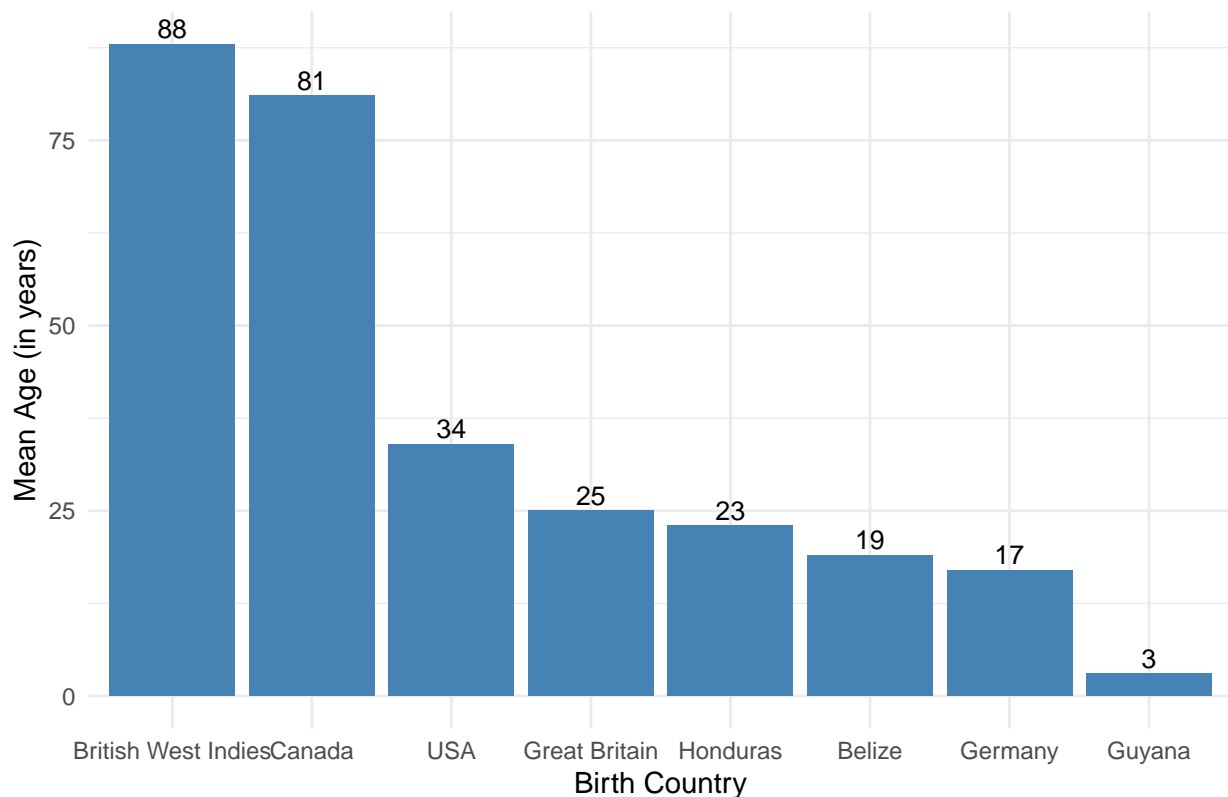
```
# bar plot of mean age of residents by their Jonestown Residence type
jonestown %>%
  group_by(genRes) %>%
  summarise(mean_age = round(mean(Age, na.rm = TRUE), digits = 0)) %>%
  ggplot(aes(x=reorder(genRes, -mean_age), y=mean_age)) +
    geom_bar(stat = "identity", fill = "steelblue")+
    geom_text(aes(label=mean_age), vjust=-0.3, size=3.5)+
    theme_minimal() +
    ggtitle("Mean Age of Residents per Residence Type") +
    ylab("Mean Age (in years)") +
    xlab("Jonestown Residence Type")
```



Mean Age of Residents by their Birth Country

```
ageByPOB # bar plot made earlier in analysis
```


Mean Age of Jonestown Residents by their Birth Country



Mean Age of US Residents by their Birth State

```
# bar plot of mean age of US residents by birth state
ageByState <- jonestown %>%
  group_by(BirthState) %>%
  summarise(mean_age = round(mean(Age, na.rm = TRUE), digits = 0)) %>%
  ggplot(aes(x=reorder(BirthState, -mean_age), y=mean_age)) +
    geom_bar(stat = "identity", fill = "steelblue")+
    geom_text(aes(label=mean_age), vjust=-0.3, size=3.5)+
    theme_minimal() +
    ggtitle("Mean Age of Jonestown Residents by their Birth State") +
    ylab("Mean Age (in years)") +
    xlab("Birth State") +
    theme(axis.text.x = element_text(angle = 90))
ageByState
```

Mean Age of Jonestown Residents by their Birth State

