

Lipids Exploration

Nick Strayer

April 4, 2016

First we read data in

```
data <- read.csv("/Users/Nick/Dropbox/vandy/regression/bios6312_final/data/ivfedata-2013_full.csv")
```

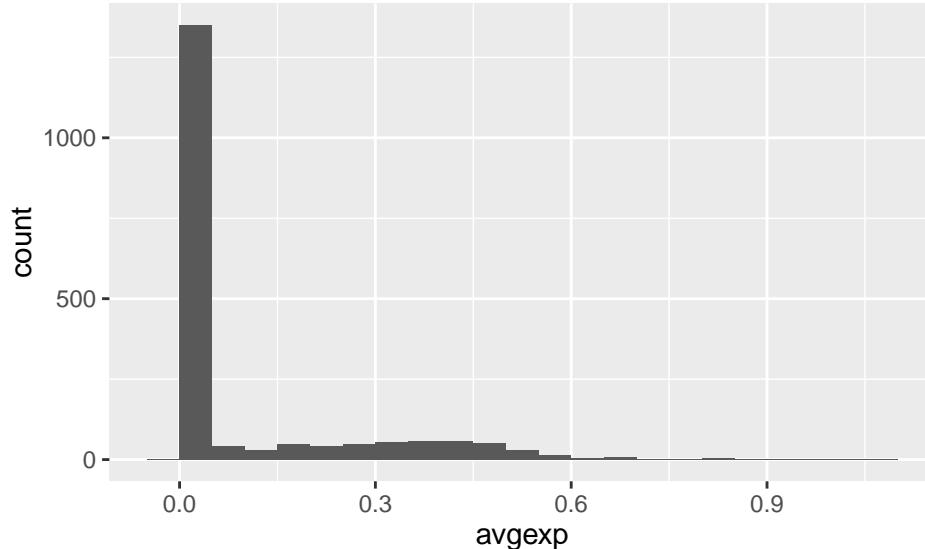
Now we recode the columns whos values are simply Ys and Ns to 1 and 0s respectively.

```
#y n columns
y_n_cols <- c("hosp.death", "unit.death", "bsi.inf", "eent.inf", "gi.inf", "lri.inf", "pneu.inf", "ssi.inf", "vital.inf")

#recode y and n values to 0=N, 1 = Y.
data[, y_n_cols] = ifelse(data[, y_n_cols] == "N", 0, 1)
```

We have a large portion of lipids users who have no exposure.

```
ggplot(data, aes(avgexp)) + geom_histogram(binwidth = 0.05)
```

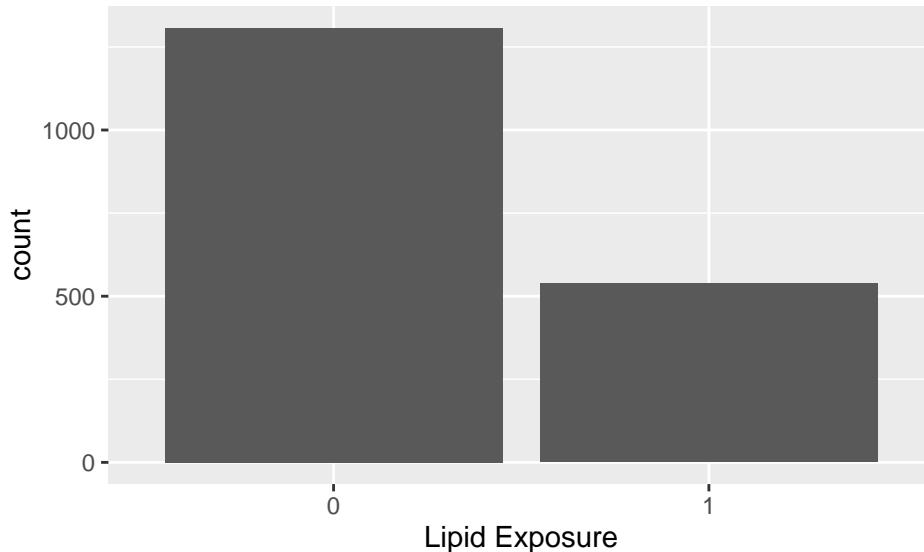


Because of this we will dichotomize on the exposure variable for logistic regression.

```
#make a binary column for if the patient received lipids during stay.
data$lipids = ifelse(data$avgexp > 0.0, 1, 0)
```

Now we check to see what the distribution of lipid use is.

```
ggplot(data, aes(x = factor(lipids))) + geom_bar(stat = "count") + labs(x = "Lipid Exposure")
```



So we can see more people did not receive them than received them.

```
kable(summary(data[, y_n_cols]))
```

hosp.death	unit.death	bsi.inf	eent.inf	gi.inf	lri.inf	pr...
Min. :0.0000	Min. :0.00000	Min. :0.00000	Min. :0.000000	Min. :0.00000	Min. :0.000000	Min. :0.000000
1st Qu.:0.0000	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.000000	1st Qu.:0.00000	1st Qu.:0.000000	1st Qu.:0.000000
Median :0.0000	Median :0.00000	Median :0.00000	Median :0.000000	Median :0.00000	Median :0.000000	Median :0.000000
Mean :0.0878	Mean :0.05691	Mean :0.02927	Mean :0.000542	Mean :0.02114	Mean :0.007588	Mean :0.000000
3rd Qu.:0.0000	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.000000	3rd Qu.:0.00000	3rd Qu.:0.000000	3rd Qu.:0.000000
Max. :1.0000	Max. :1.00000	Max. :1.00000	Max. :1.000000	Max. :1.00000	Max. :1.00000	Max. :1.000000

Descriptive Statistics of Patients.

First a quick function to generate a nice table output of `summary`.

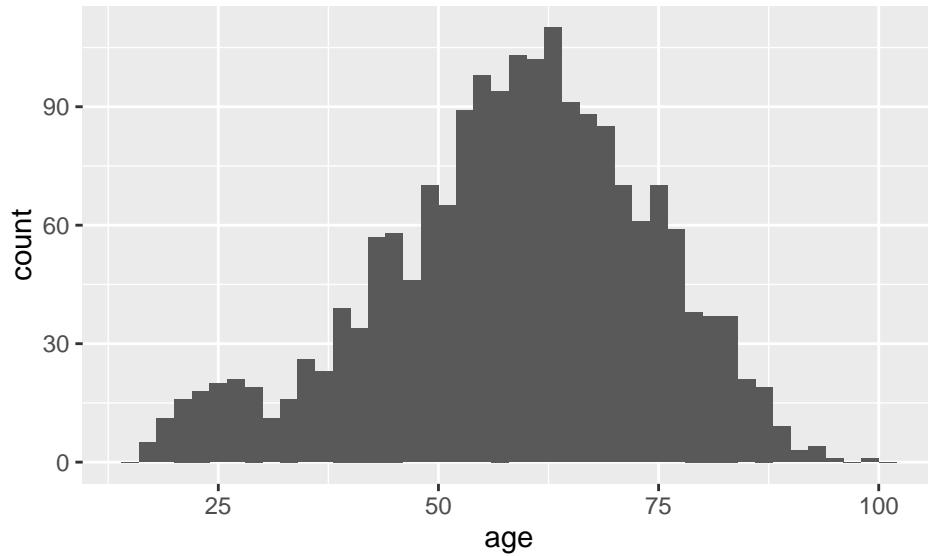
```
#this is tedious
summaryTable <- function(variable){ kable(as.data.frame(as.list(summary(variable)))) }
```

Ages

```
summaryTable(data$age)
```

Min.	X1st.Qu.	Median	Mean	X3rd.Qu.	Max.
17.02	49.4	59.66	58.47	69.08	98.64

```
ggplot(data, aes(age)) + geom_histogram(binwidth = 2) + labs(x = "age")
```

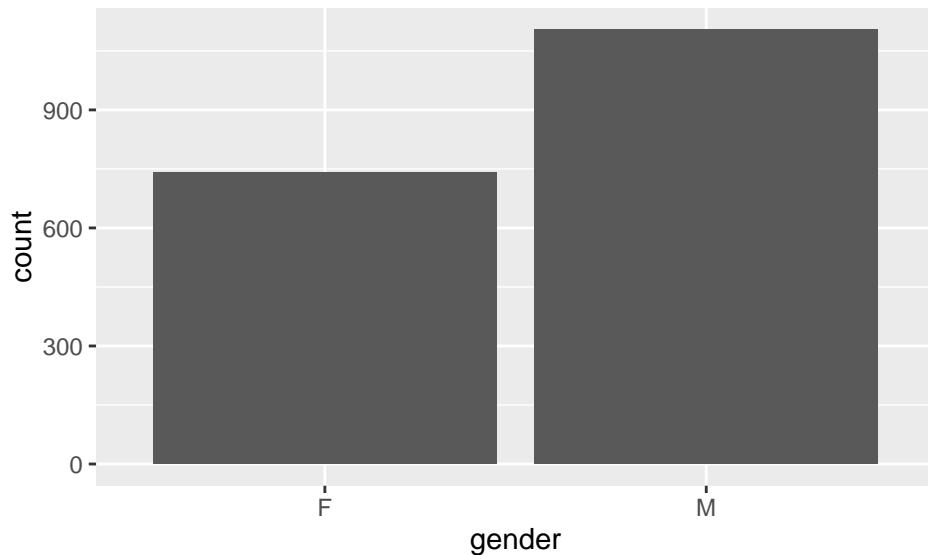


Gender

```
summaryTable(data$gender)
```

	F	M
	741	1104

```
ggplot(data, aes(x = factor(gender))) + geom_bar(stat = "count") + labs(x = "gender")
```

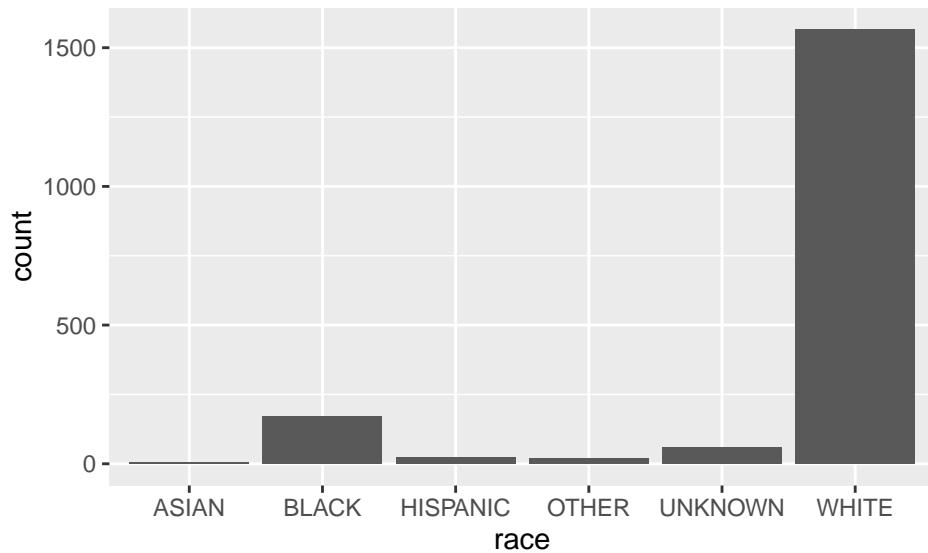


Race

```
#recode the ? race to unknown.  
data[data$race == "?", "race"] = "UNKNOWN"  
  
summaryTable(data$race)
```

X.	ASIAN	BLACK	HISPANIC	OTHER	UNKNOWN	WHITE
0	6	173	22	19	59	1566

```
ggplot(data, aes(x = factor(race))) + geom_bar(stat = "count") + labs(x = "race")
```

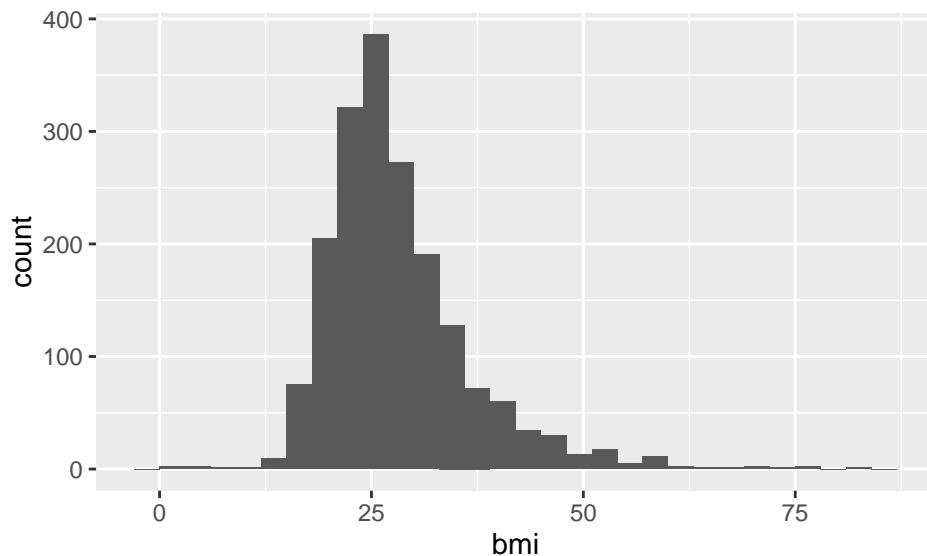


BMI

```
summaryTable(data$bmi)
```

Min.	X1st.Qu.	Median	Mean	X3rd.Qu.	Max.
0.15	22.61	26.26	28.09	31.43	83.12

```
ggplot(data, aes(bmi)) + geom_histogram(binwidth = 3)
```

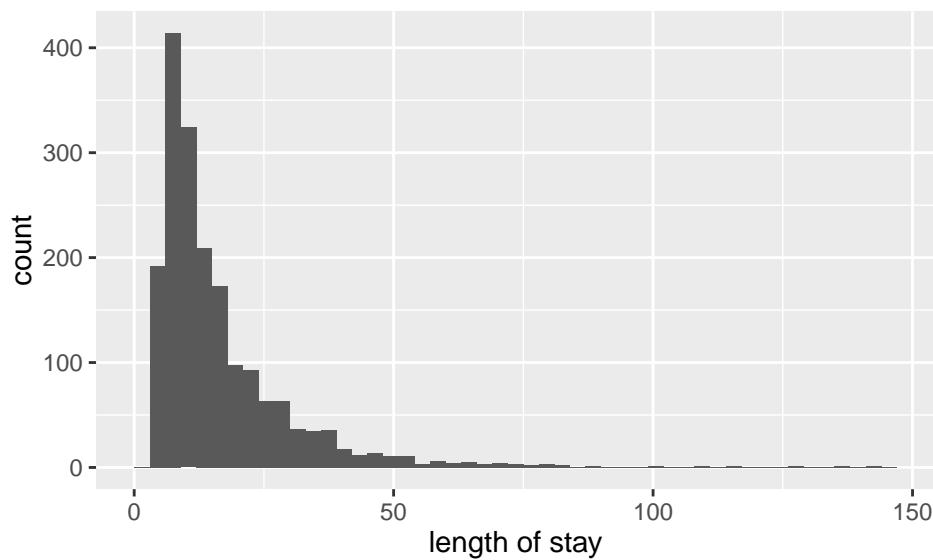


Hospital Length of Stay

```
summaryTable(data$hosp.los)
```

Min.	X1st.Qu.	Median	Mean	X3rd.Qu.	Max.
3.003	8.005	11.87	16.41	20.24	141.4

```
ggplot(data, aes(hosp.los)) + geom_histogram(binwidth = 3) + labs(x = "length of stay")
```

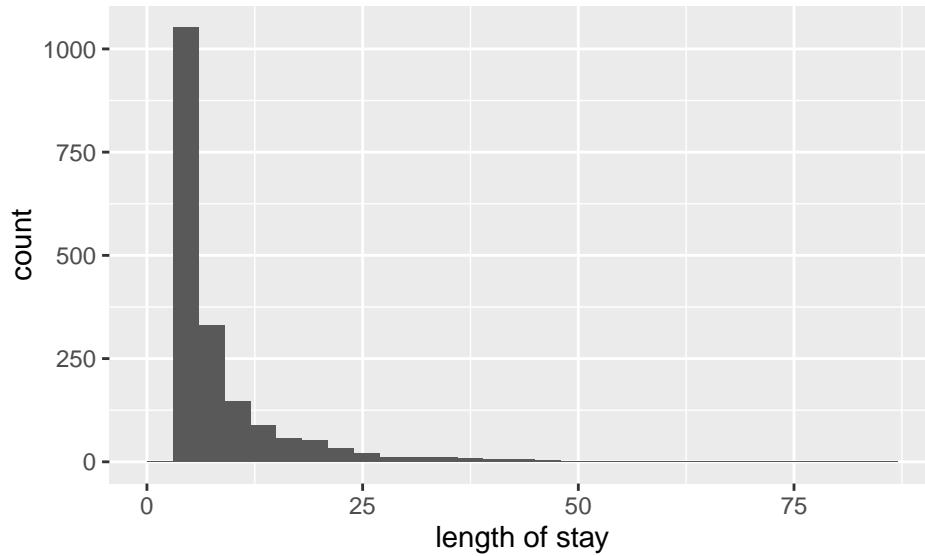


Unit Length of Stay

```
summaryTable(data$unit.los)
```

Min.	X1st.Qu.	Median	Mean	X3rd.Qu.	Max.
3.001	3.898	5.265	8.287	9.024	82.51

```
ggplot(data, aes(unit.los)) + geom_histogram(binwidth = 3) + labs(x = "length of stay")
```

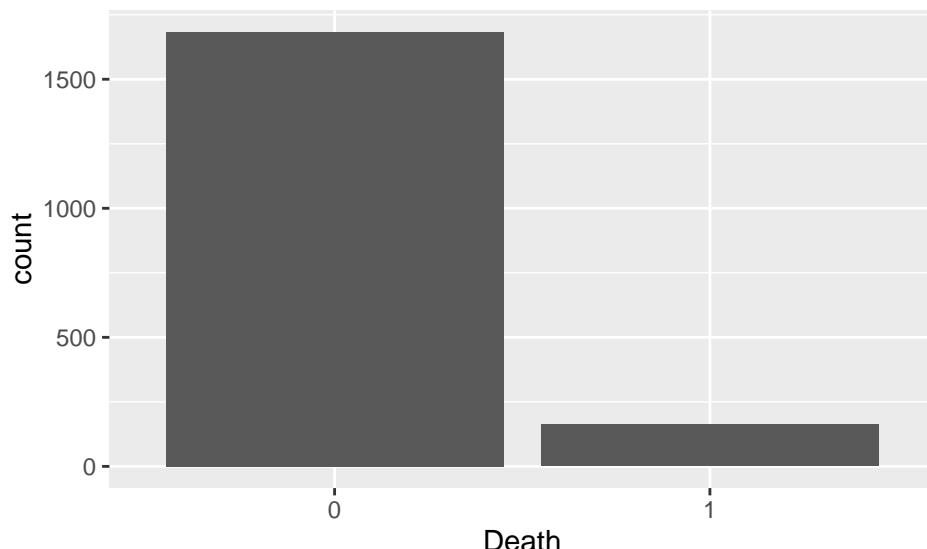


Hospital Deaths

```
summaryTable(data$hosp.death)
```

Min.	X1st.Qu.	Median	Mean	X3rd.Qu.	Max.
0	0	0	0.0878	0	1

```
ggplot(data, aes(x = factor(hosp.death))) + geom_bar(stat = "count") + labs(x = "Death")
```



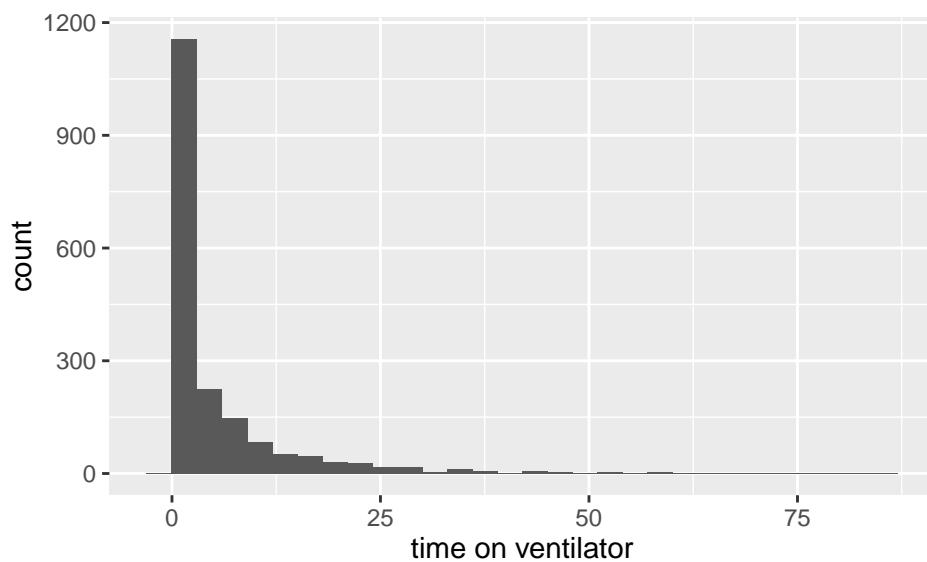
A much greater amount of people didn't die.

Time on Ventilator

```
summaryTable(data$ventdays.unit)
```

Min.	X1st.Qu.	Median	Mean	X3rd.Qu.	Max.
0	0	1.328	4.878	6	82.51

```
ggplot(data, aes(ventdays.unit)) + geom_histogram(binwidth = 3) + labs(x = "time on ventilator")
```

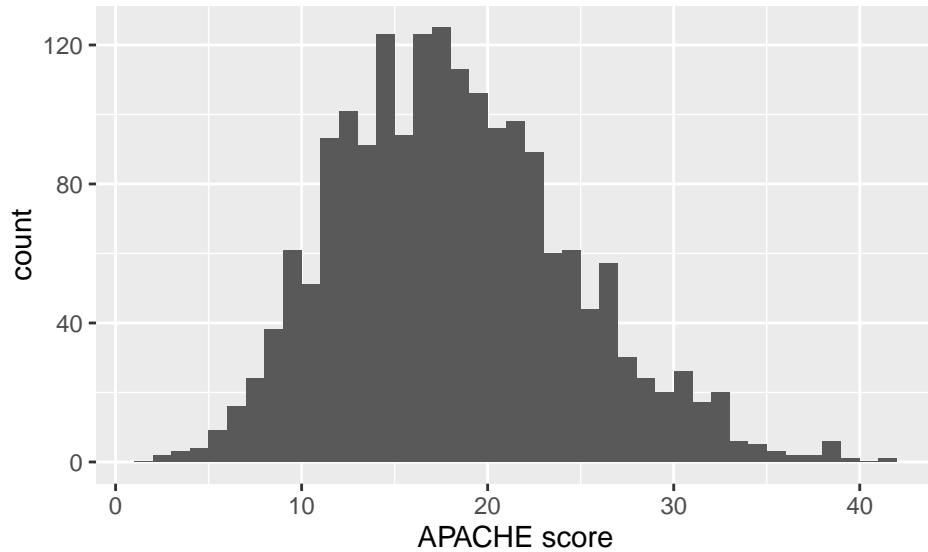


Apache Score.

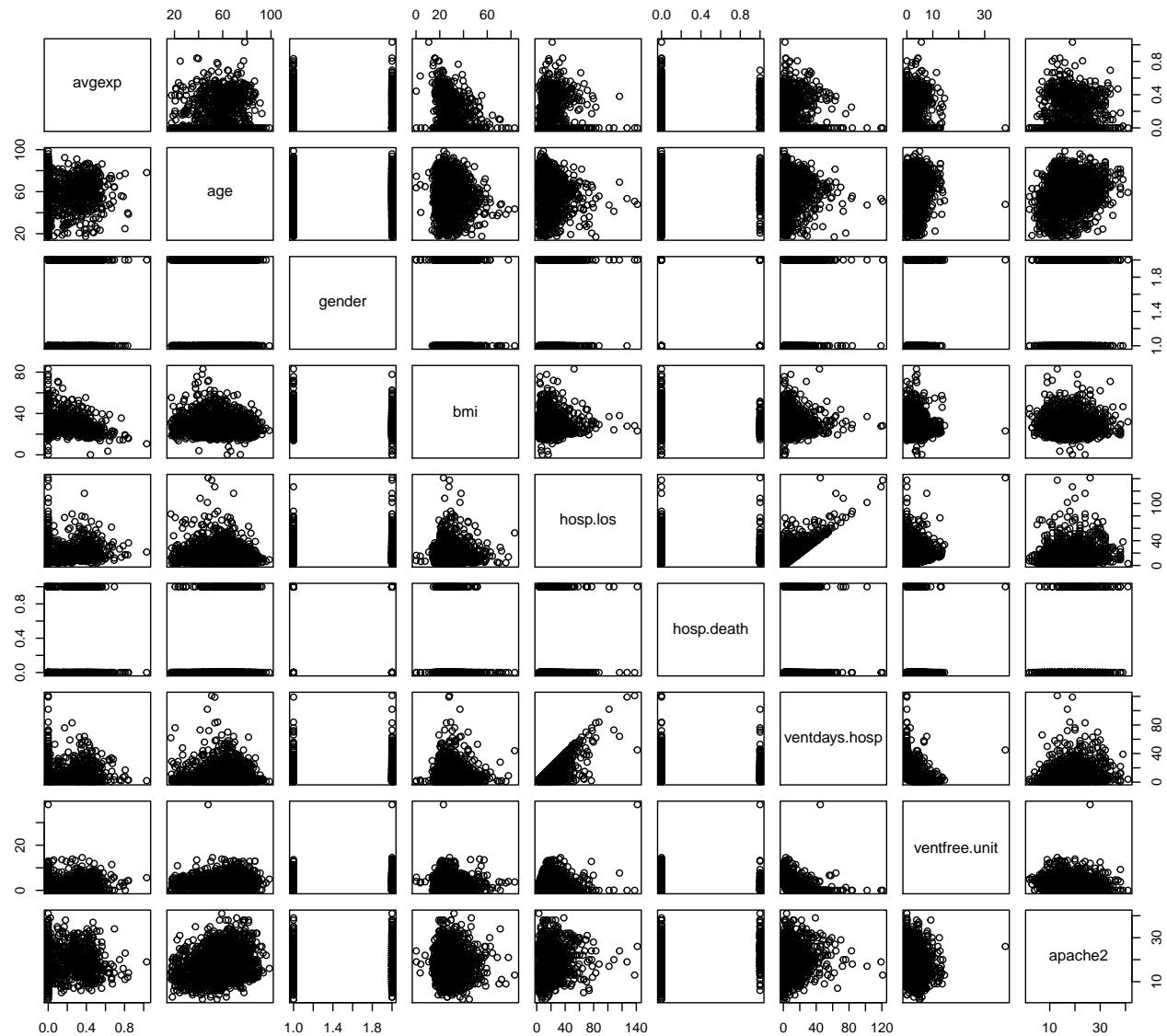
```
summaryTable(data$ventdays.apache2)
```

	Length	Class	Mode
0		NULL	NULL

```
ggplot(data, aes(apache2)) + geom_histogram(binwidth = 1) + labs(x = "APACHE score")
```



```
demoCols <- c("avgexp", "age", "gender", "bmi", "hosp.los", "hosp.death", "ventdays.hosp",  
plot(data[,demoCols])
```



```

controls = c("bmi", "glucose", "apache2", "race", "gender", "age")

plot(data[, controls])

summary(lm(data$apache2 ~ data$age))

table(data$apache2)

```