

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")
```

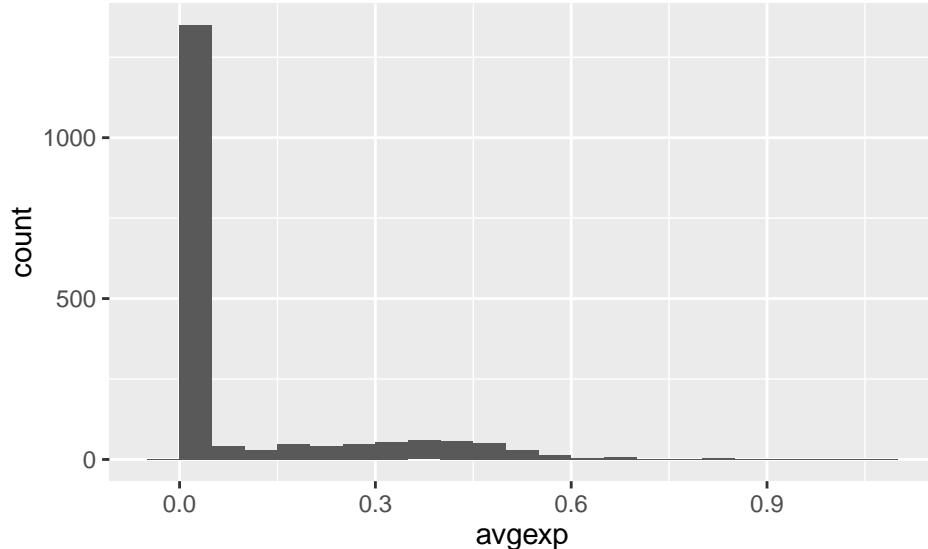
Now we recode the columns whos values are simply Ys and Ns to 1 and 0s respectively.

```
#y n columns
y_n_cols <- c("hosp.death", "unit.death", "bsi.inf", "eent.inf", "gi.inf",
             "lri.inf", "pneu.inf", "ssi.inf", "sst.inf", "sys.inf", "uti.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)
y_n_cols = c("lipids", y_n_cols) #add this to the dichotomous variables now.
```

Continuous Variables

Table of Descriptive Statistics

```
continuousVariables1 <- c("avgexp", "maxexp", "age", "bmi", "glucose", "hosp.los", "unit.los")
continuousVariables2 <- c("ventdays.hosp", "ventdays.unit", "ventfree.unit", "apache2")

options(digits=2)
kable(stat.desc(data[, continuousVariables1], basic=F))
```

	avgexp	maxexp	age	bmi	glucose	hosp.los	unit.los
median	0.00	0.00	59.66	26.26	170.00	11.87	5.26
mean	0.09	0.15	58.47	28.09	184.14	16.41	8.29
SE.mean	0.00	0.01	0.36	0.20	1.58	0.32	0.19
CI.mean.0.95	0.01	0.01	0.70	0.38	3.10	0.63	0.37
var	0.03	0.08	236.44	70.87	4601.82	190.79	64.29
std.dev	0.17	0.29	15.38	8.42	67.84	13.81	8.02
coef.var	1.89	1.97	0.26	0.30	0.37	0.84	0.97

```
kable(stat.desc(data[, continuousVariables2], basic=F))
```

	ventdays.hosp	ventdays.unit	ventfree.unit	apache2
median	3.00	1.33	3.42	17.00
mean	7.59	4.88	3.41	17.72
SE.mean	0.27	0.20	0.06	0.15
CI.mean.0.95	0.53	0.39	0.12	0.29
var	134.20	73.65	6.71	39.47
std.dev	11.58	8.58	2.59	6.28
coef.var	1.53	1.76	0.76	0.35

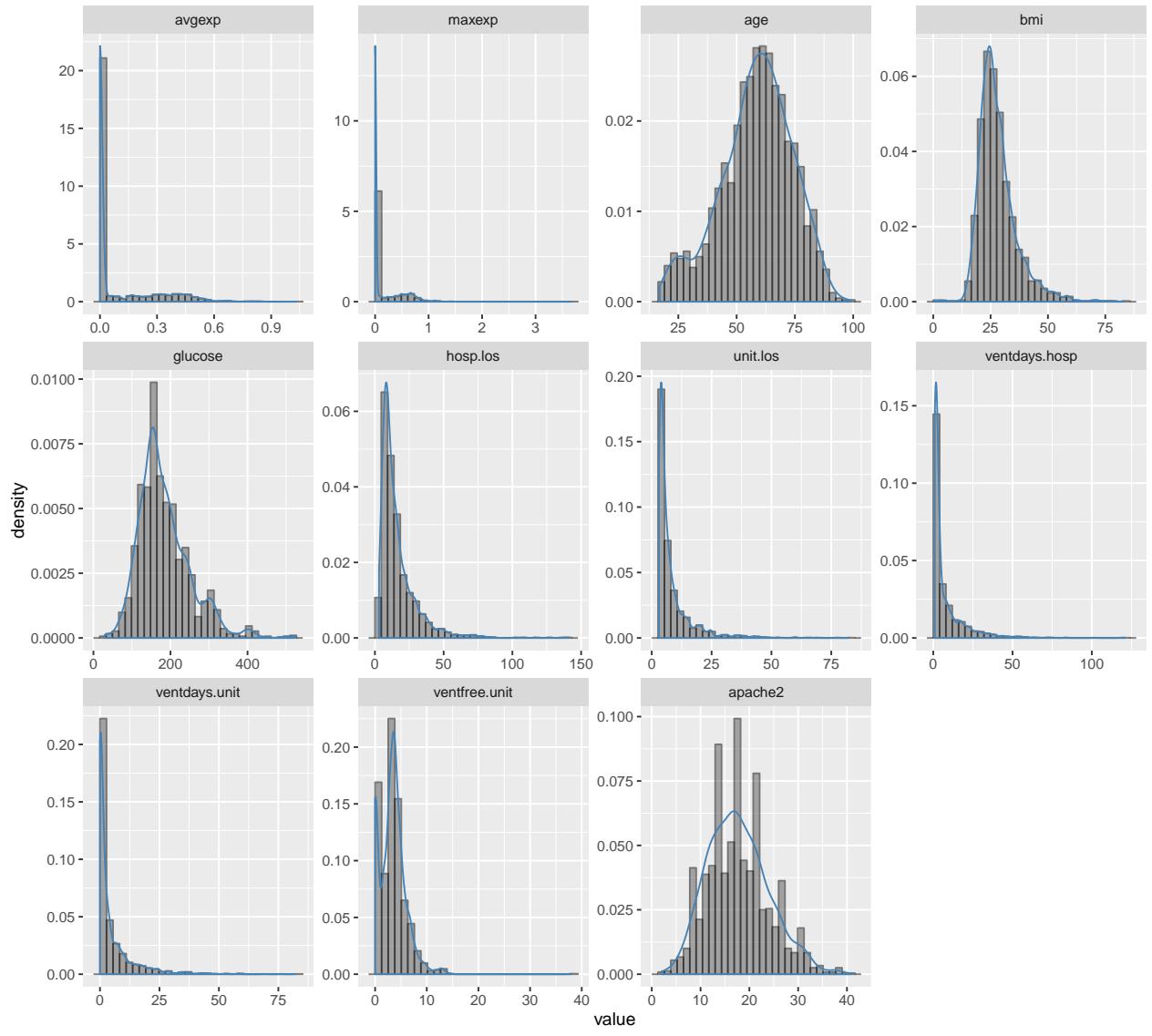
Plot of histograms

```
cvs <- c("avgexp", "maxexp", "age", "bmi", "glucose", "hosp.los", "unit.los",
        "ventdays.hosp", "ventdays.unit", "ventfree.unit", "apache2")
cvplot <- data[, cvs]
cvplot <- melt(cvplot)
```

```
## Using as id variables
```

```
ggplot(cvplot, aes(value)) +
  geom_histogram(aes(y = ..density..), col="black", alpha = 0.5) +
  facet_wrap(~variable, scales = "free") +
  geom_density(color = "steelblue")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Dichotomous Variables

Outcomes plot:

```
tp <- melt(data[, y_n_cols])
```

```
## Using   as id variables
```

```
ggplot(data.frame(table(tp)), aes(x = 1, y = Freq, fill = value)) +  
  geom_bar(stat="identity", position=position_stack()) +  
  facet_wrap(~ variable) + labs(x = "")
```



```
stats <- data.frame( no_occur = integer(), occur = integer())  
for(i in 1:length(y_n_cols)) stats[i,] <- table(data[, y_n_cols[i]])  
rownames(stats) <- c("Lipid Emulsion", "Death in Hospital", "Death in Unit", "bsi.inf", "eent.inf", "gi.inf",  
kable(stats, col.names= c("Occured", "Did not Occur"))
```

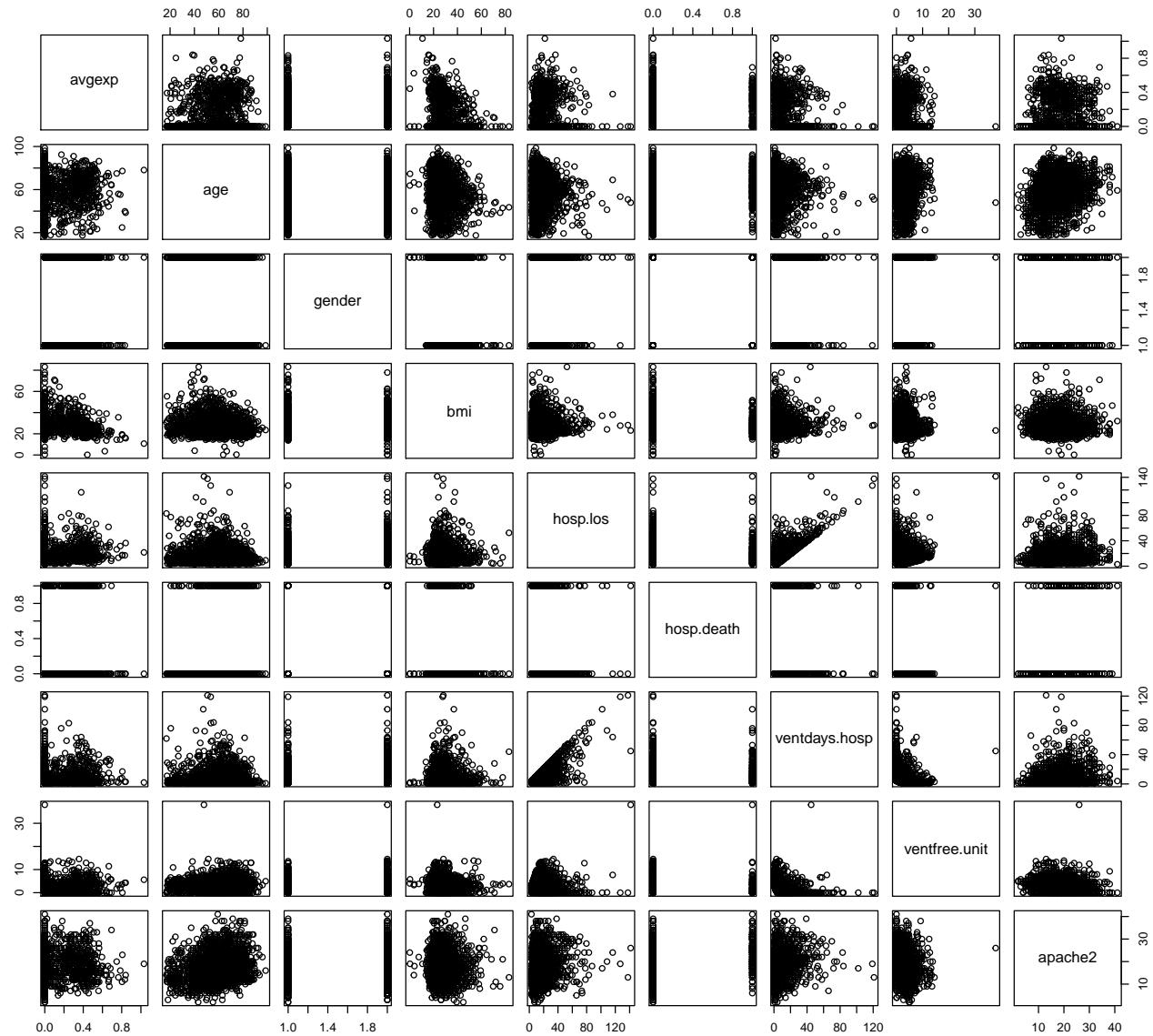
	Occured	Did not Occur
Lipid Emulsion	1307	538

	Occured	Did not Occur
Death in Hospital	1683	162
Death in Unit	1740	105
bsi.inf	1791	54
eent.inf	1844	1
gi.inf	1806	39
lri.inf	1831	14
pneu.inf	1706	139
ssi.inf	1796	49
sst.inf	1844	1
sys.inf	1845	1845
uti.inf	1806	39

Big Scatterplots

Looking at relationships between variables.

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



Same thing but with the variables we chose as controls in our regressions.

```
controls = c("bmi", "glucose", "apache2", "race", "gender", "age")  
plot(data[, controls])
```

