pcancer distribution

Don Ghysels

2/21/2021

```r
library('readxl')
library("rpart")
library("corrplot")

pcancerdf <- read_excel("../pcancer.xlsx")

pcancer.cor <- cor(pcancerdf[,2:9])

pcancerdf.psa.reduced <- pcancerdf[pcancerdf$psa < 200,]
pcancer.cor.psa <- cor(pcancerdf.psa.reduced[,2:9])
cordf <- data.frame(pcancer.cor[1,],pcancer.cor.psa[1,],pcancer.cor.psa[1,] - pcancer.cor[1,])
colnames(cordf) <-  c("cor","reduced cor", "improvement")
knitr::kable(cordf)
```
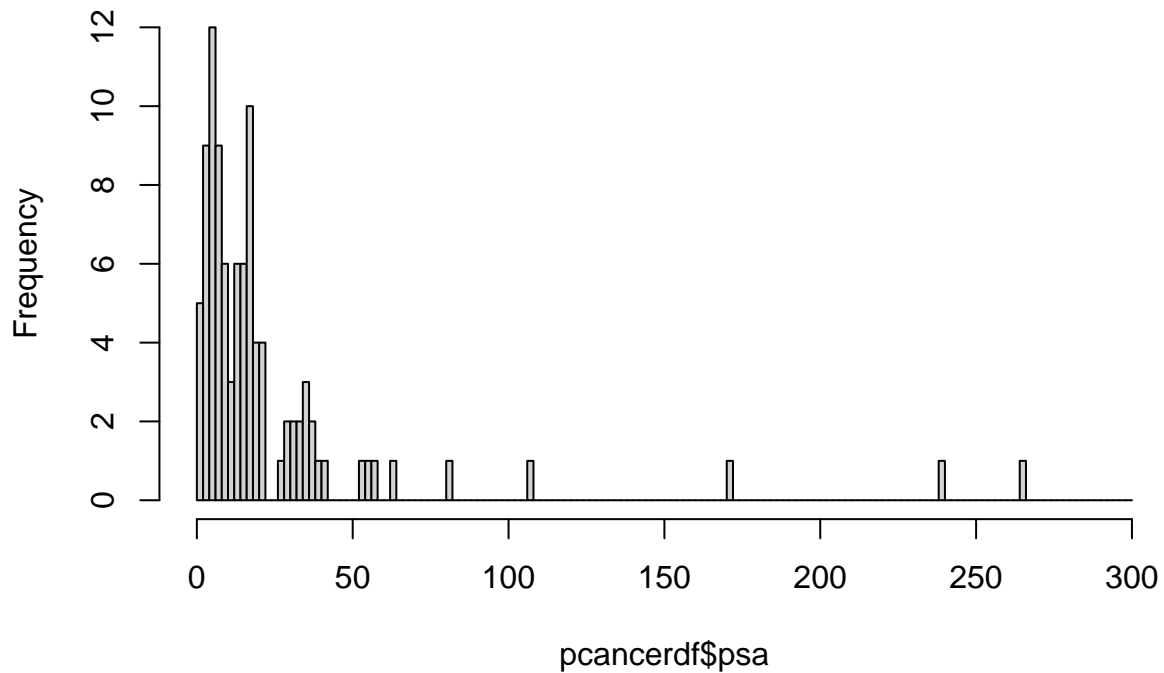
	cor	reduced cor	improvement
psa	1.0000000	1.0000000	0.0000000
cancerv	0.6241506	0.6463295	0.0221789
weight	0.0262134	0.0322836	0.0060701
age	0.0171994	-0.0858012	-0.1030005
hyperplasia	-0.0164865	-0.0656469	-0.0491604
seminal	0.5286188	0.5493860	0.0207672
capsular	0.5507925	0.4821814	-0.0686111
score	0.4295798	0.4440317	0.0144519

```r
hist(pcancerdf$psa, breaks = seq(0,300, 2))
```

## Histogram of pcancerdf$psa


```
pcancerdf.cv.reduced <- pcancerdf[pcancerdf$cancerv < 30,]
pcancer.cor.cv <- cor(pcancerdf.cv.reduced[,2:9])
cordf <- data.frame(pcancer.cor[2,],pcancer.cor.cv[2,],pcancer.cor.cv[2,] - pcancer.cor[2,])
colnames(cordf) <-  c("cor","reduced cor", "improvement")
knitr::kable(cordf)
```

	cor	reduced cor	improvement
psa	0.6241506	0.5169959	-0.1071547
cancerv	1.0000000	1.0000000	0.0000000
weight	0.0051071	-0.0061761	-0.0112832
age	0.0390944	0.2070564	0.1679620
hyperplasia	-0.1332094	-0.0990966	0.0341129
seminal	0.5817417	0.5461632	-0.0355785
capsular	0.6928967	0.6522521	-0.0406446
score	0.4814384	0.4506972	-0.0307412

```
hist(pcancerdf$cancerv, breaks = seq(0,50, 0.5))
```
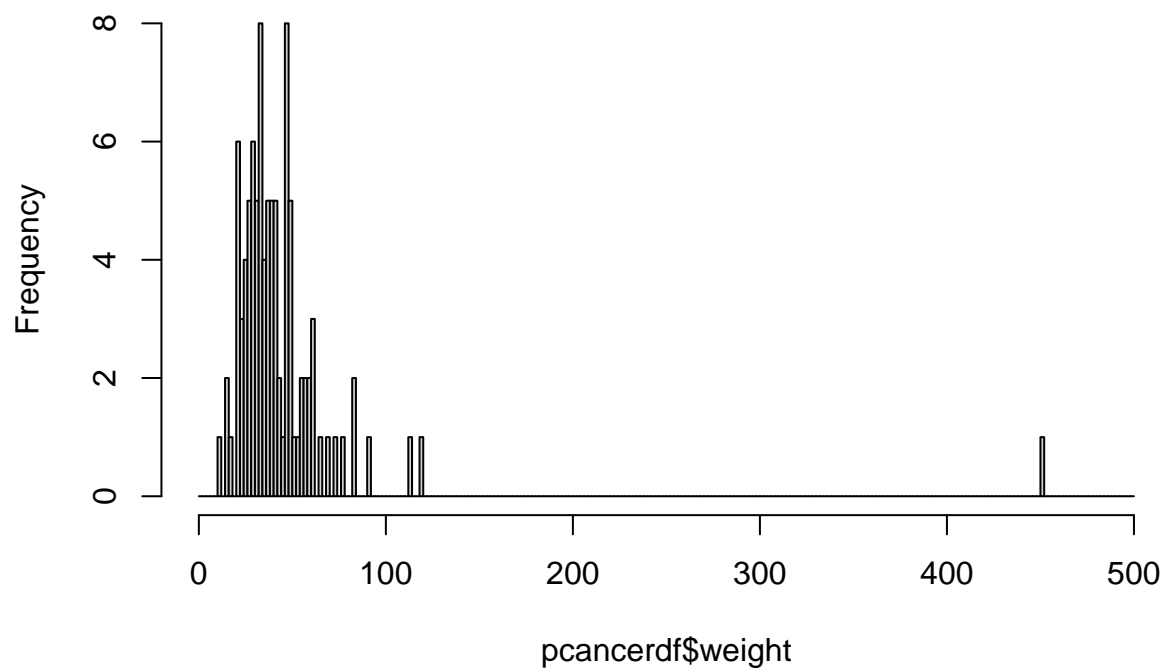
Histogram of pcancerdf$cancerv
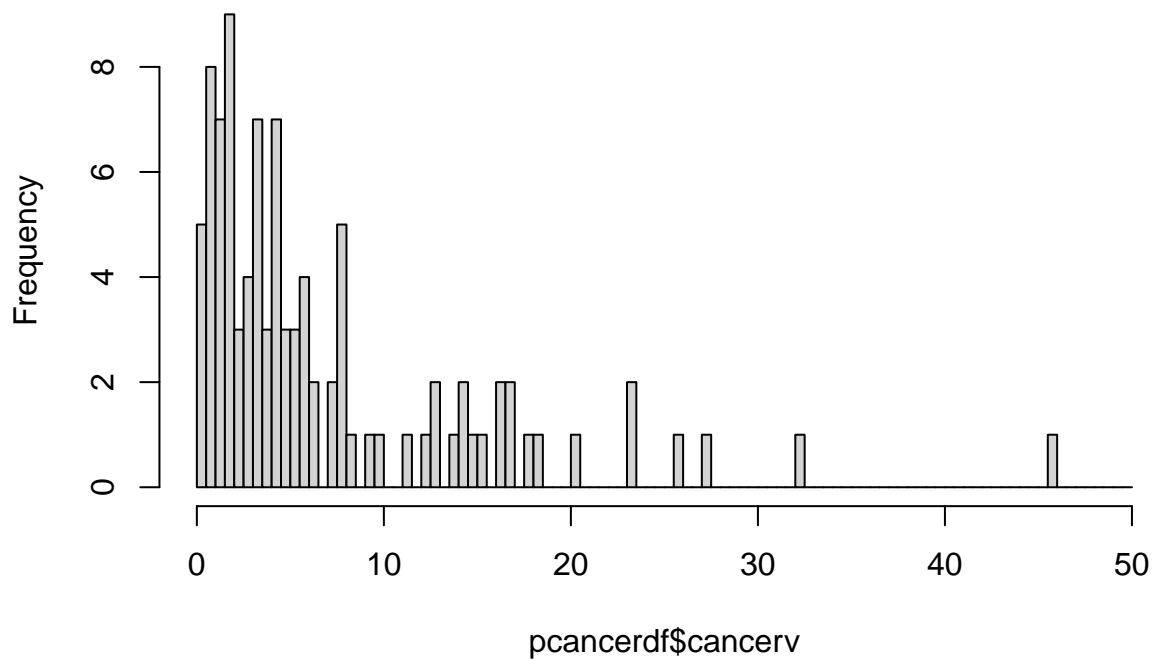

```
pcancerdf.wt.reduced <- pcancerdf[pcancerdf$weight < 400,]
pcancer.cor.wt <- cor(pcancerdf.wt.reduced[,2:9])
cordf <- data.frame(pcancer.cor[3,],pcancer.cor.wt[3,],pcancer.cor.wt[3,] - pcancer.cor[3,])
colnames(cordf) <-  c("cor","reduced cor", "improvement")
knitr::kable(cordf)
```

	cor	reduced cor	improvement
psa	0.0262134	0.1520382	0.1258248
cancerv	0.0051071	0.1771875	0.1720804
weight	1.0000000	1.0000000	0.0000000
age	0.1643237	0.3597479	0.1954242
hyperplasia	0.3218487	0.5573283	0.2354795
seminal	-0.0024105	0.1112017	0.1136122
capsular	0.0015789	0.1366134	0.1350345
score	-0.0242069	0.2082057	0.2324126

```
hist(pcancerdf$weight, breaks = seq(0,500, 2))
```

Histogram of pcancerdf$weight
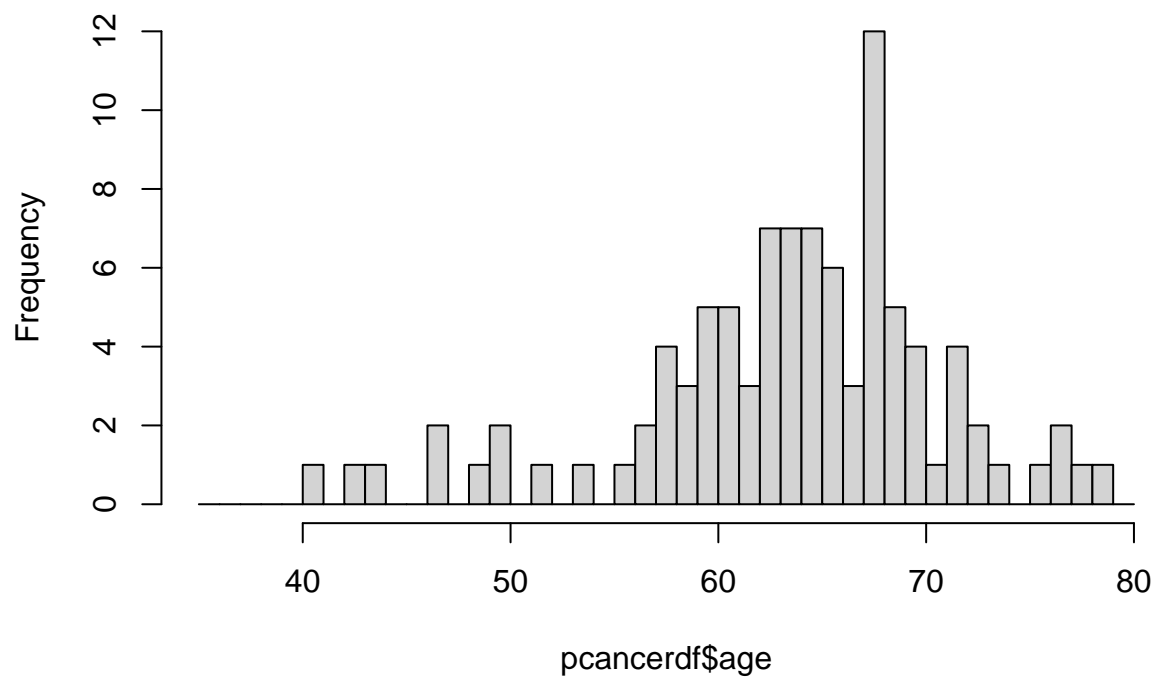


```
pcancer.cor[4,]
```

```
##        psa      cancerv      weight          age hyperplasia     seminal
##  0.01719938  0.03909442  0.16432371  1.00000000  0.36634121  0.11765804
##    capsular        score
##  0.09955535  0.22585181
```

```
hist(pcancerdf$age, breaks = seq(35,80, 1))
```

Histogram of pcancerdf$age
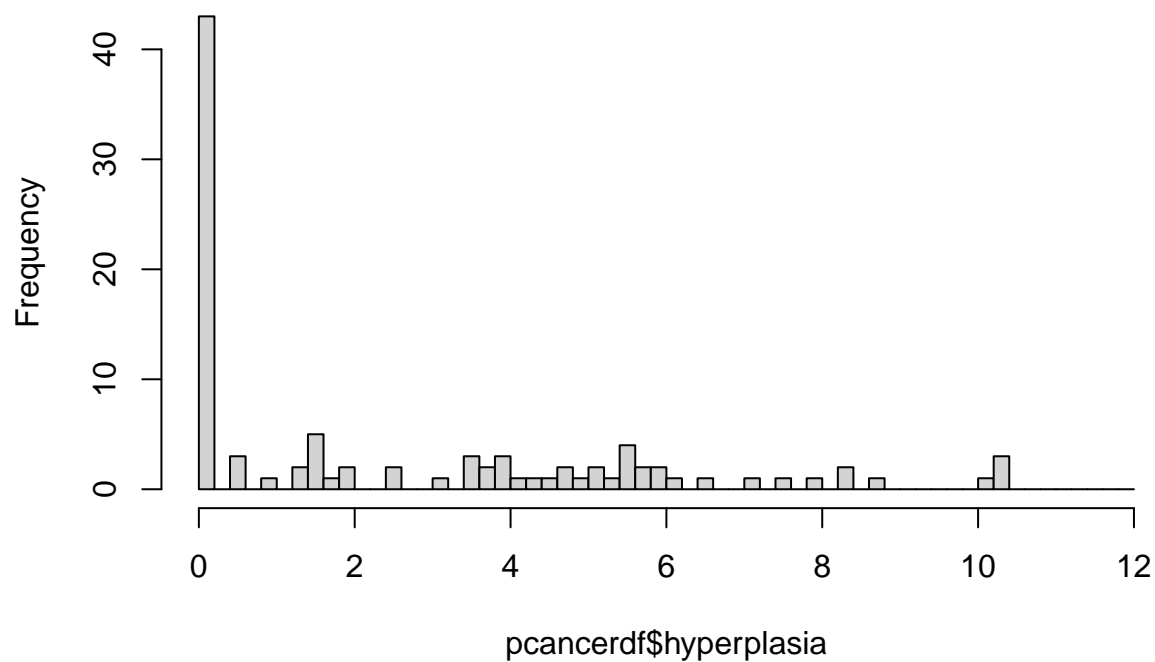


```
pcancer.cor[5,]
```

```
##         psa     cancerv      weight         age hyperplasia      seminal
## -0.01648649 -0.13320943  0.32184875  0.36634121  1.00000000 -0.11955319
##    capsular       score
## -0.08300865  0.02682555
```

```
hist(pcancerdf$hyperplasia, breaks = seq(0,12, 0.2))
```

Histogram of pcancerdf$hyperplasia
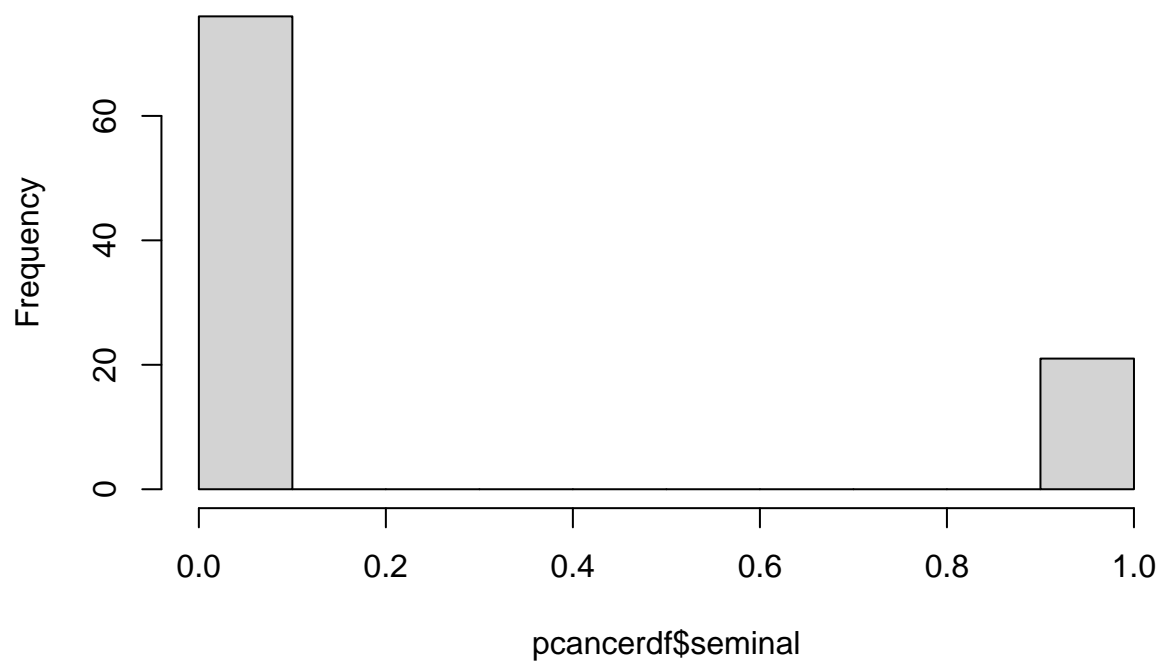


```
pcancer.cor[6,]
```

```
##         psa      cancerv       weight         age  hyperplasia       seminal
##  0.528618785  0.581741687 -0.002410475  0.117658038 -0.119553192  1.000000000
##    capsular        score
##  0.680284092  0.428573479
```

```
hist(pcancerdf$seminal, breaks = seq(0,1, .1))
```

Histogram of pcancerdf$seminal
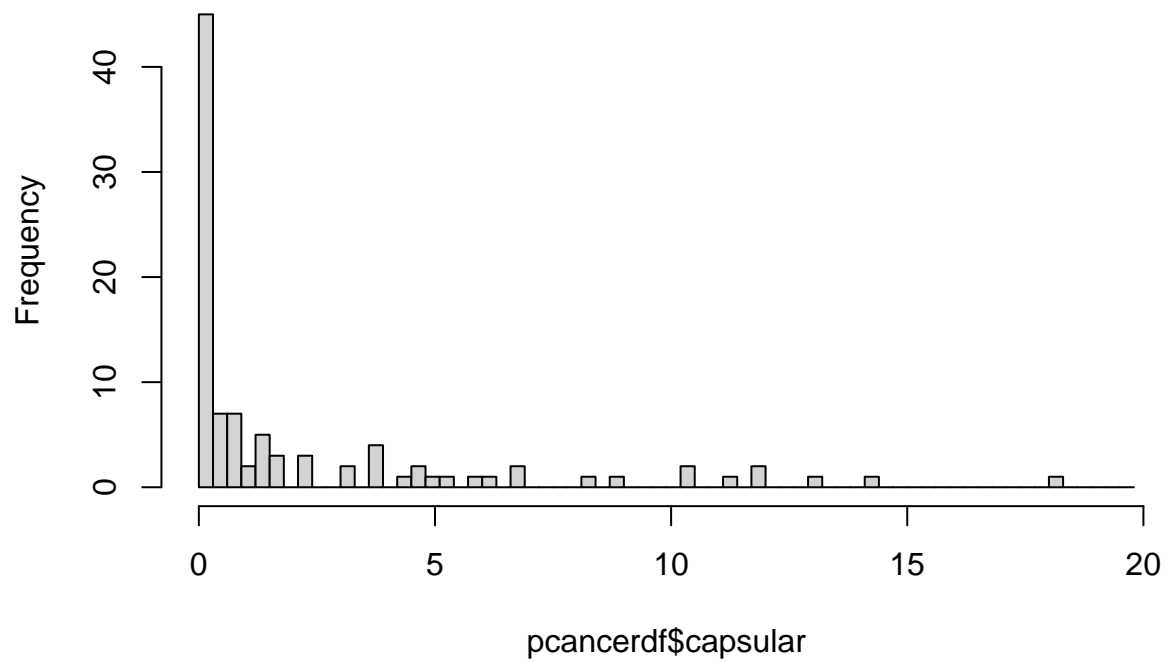


```
pcancer.cor[7,]
```

```
##          psa      cancerv       weight          age  hyperplasia      seminal
##   0.550792517  0.692896688  0.001578905  0.099555351 -0.083008649  0.680284092
##      capsular        score
##   1.000000000  0.461565896
```

```
hist(pcancerdf$capsular, breaks = seq(0,20, 0.3))
```

Histogram of pcancerdf$capsular


```
pcancer.cor[8,]
```

```
##         psa      cancerv       weight          age hyperplasia      seminal
##  0.42957975   0.48143840  -0.02420693   0.22585181   0.02682555   0.42857348
##     capsular        score
##  0.46156590   1.00000000
```

```
hist(pcancerdf$score, breaks = seq(5,8,0.1))
```

Histogram of pcancerdf$score