Data Analysis

Juste Simanauskaite & Patricia Rivera

Some test words here

```
knitr::opts_chunk$set(message=FALSE, warning=FALSE, fig.height=3, fig.width=5, fig.align="center")
library(tidyverse)
library(broom)
library(plyr)
library(survival)
library(survminer)
aids <- read.csv( "http://pages.pomona.edu/~jsh04747/courses/math150/AIDSdata.csv")
dim(aids)
```

```
## [1] 851  16
```

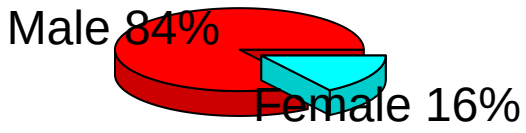
```
summary(aids)
```

```
##        id              time           censor            time_d
##  Min.   :   1.0   Min.   :  1.0   Min.   :0.00000   Min.   :  1.0
##  1st Qu.: 287.5   1st Qu.:179.5   1st Qu.:0.00000   1st Qu.:199.5
##  Median : 581.0   Median :257.0   Median :0.00000   Median :266.0
##  Mean   : 579.5   Mean   :231.8   Mean   :0.08108   Mean   :243.4
##  3rd Qu.: 873.0   3rd Qu.:300.0   3rd Qu.:0.00000   3rd Qu.:306.0
##  Max.   :1156.0   Max.   :362.0   Max.   :1.00000   Max.   :362.0
##     censor_d            tx              txgrp           strat2
##  Min.   :0.0000   Min.   :0.0000   Min.   :1.000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:0.0000
##  Median :0.0000   Median :1.0000   Median :2.000   Median :1.0000
##  Mean   :0.0235   Mean   :0.5041   Mean   :1.504   Mean   :0.6157
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :2.000   Max.   :1.0000
##       sex            raceth          ivdrug          hemophil
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :0.00000
##  1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:0.00000
##  Median :1.000   Median :1.000   Median :1.000   Median :0.00000
##  Mean   :1.157   Mean   :1.706   Mean   :1.317   Mean   :0.03408
##  3rd Qu.:1.000   3rd Qu.:2.000   3rd Qu.:1.000   3rd Qu.:0.00000
##  Max.   :2.000   Max.   :5.000   Max.   :3.000   Max.   :1.00000
##      karnof            cd4            priorzdv          age
##  Min.   : 70.00   Min.   :  0.00   Min.   :  3.00   Min.   :15.00
##  1st Qu.: 90.00   1st Qu.: 22.25   1st Qu.: 11.00   1st Qu.:33.00
##  Median : 90.00   Median : 75.00   Median : 21.00   Median :38.00
##  Mean   : 91.34   Mean   : 86.45   Mean   : 30.63   Mean   :38.81
##  3rd Qu.:100.00   3rd Qu.:135.75   3rd Qu.: 44.00   3rd Qu.:44.00
##  Max.   :100.00   Max.   :348.00   Max.   :288.00   Max.   :73.00
```

The data set contains a sample size equal to 851 participants and 16 different variables.

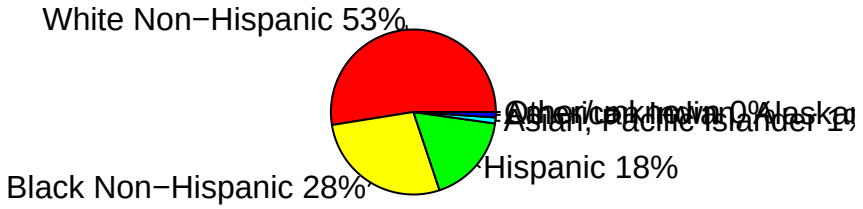
```
library(plotrix)
male<-sum(aids$sex==1)
female<-sum(aids$sex==2)
slices <- c(male, female)
lbls <- c("Male", "Female")
```

```
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct)
lbls <- paste(lbls,"%",sep="")
pie3D(slices,labels=lbls,explode=0.1,
   main="Gender Distribution ")
```

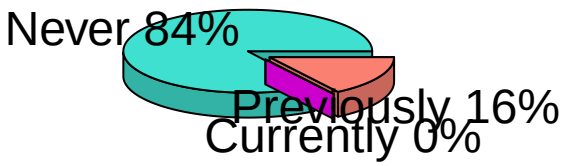


The Pie Chart represents the gender distribution in the sample, with 84% male and 16% female.

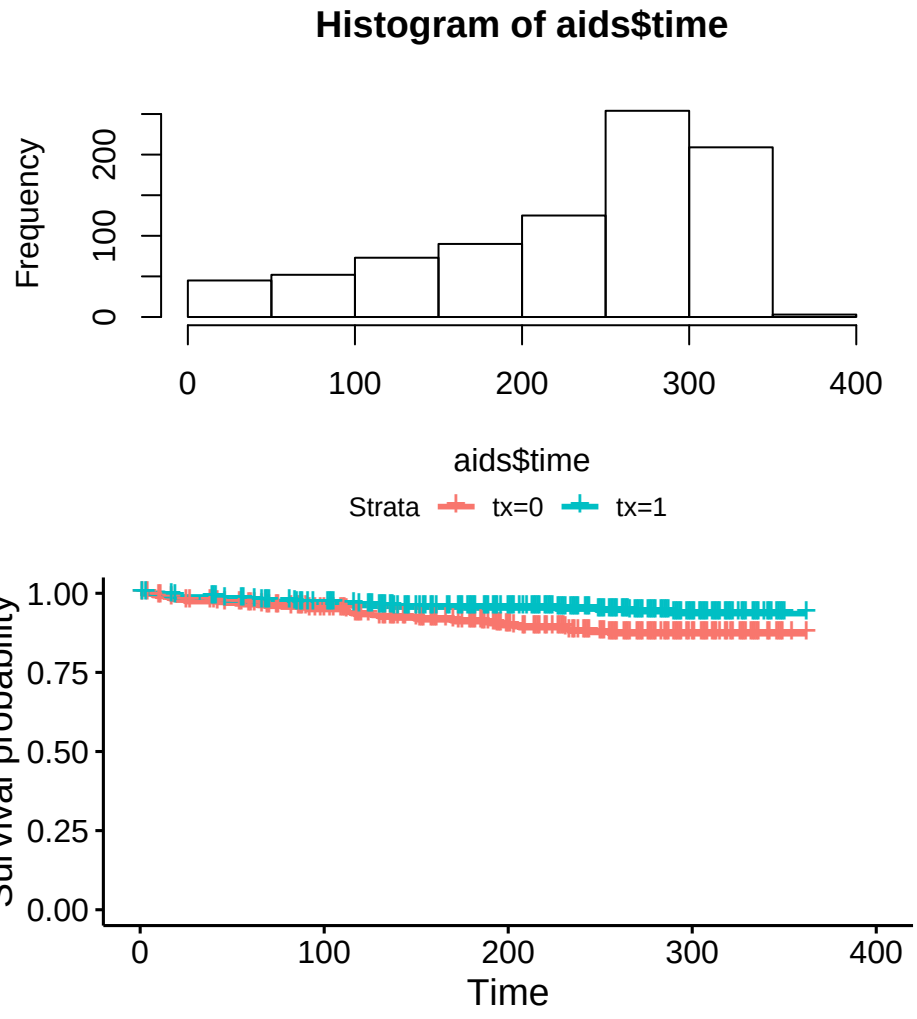
```
wnh<-sum(aids$raceth==1)
bnh<-sum(aids$raceth==2)
h<-sum(aids$raceth==3)
api<-sum(aids$raceth==4)
aian<-sum(aids$raceth==5)
oth<-sum(aids$raceth==6)
slices <- c(wnh,bnh,h,api,aian,oth)
lbls <- c("White Non-Hispanic", "Black Non-Hispanic", "Hispanic","Asian, Pacific Islander", "American In
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct)
lbls <- paste(lbls,"%",sep="")
pie(slices,lbls,col = rainbow(length(lbls)) )
```



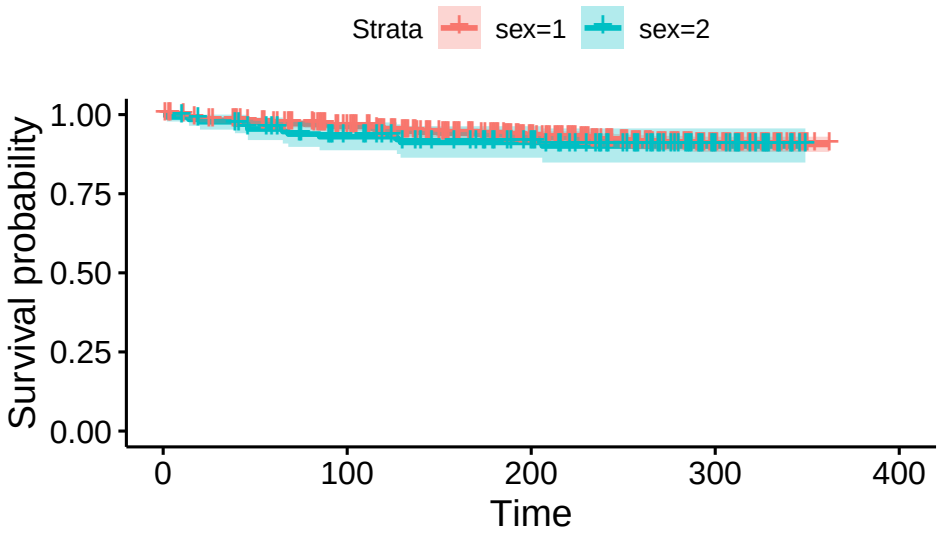

```
never<-sum(aids$ivdrug==1)
cur<-sum(aids$ivdrug==2)
prev<-sum(aids$ivdrug==3)
slices <- c(never,cur,prev)
lbls <- c("Never", "Currently", "Previously")
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct)
lbls <- paste(lbls,"%",sep="")
pie3D(slices,labels=lbls,explode=0.1,col=c("turquoise","magenta","salmon"),cex=0.5,
   main="IV Drug Use History ")
```



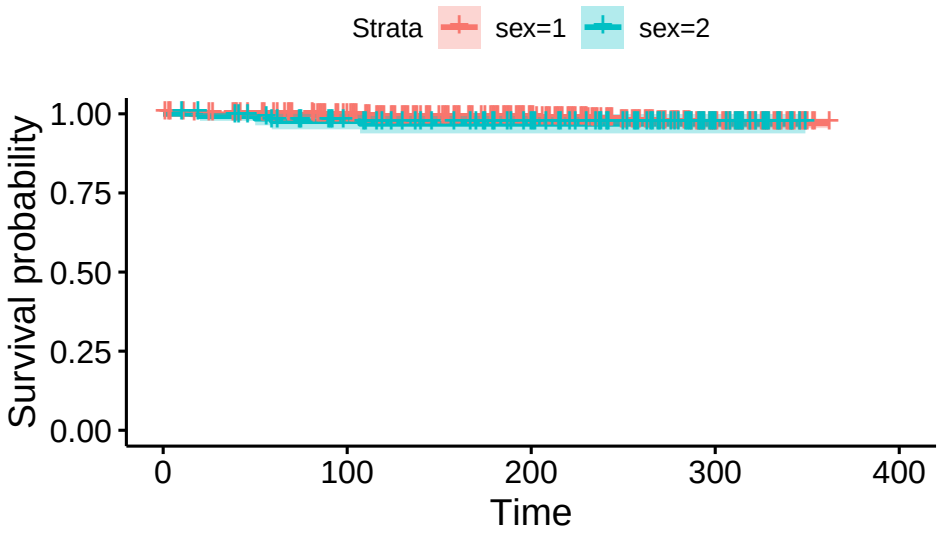

```
hist(aids$time)

###Data Plots

fit <- survfit(Surv(time,censor)~tx, data = aids)
ggsurvplot(fit,data = aids,conf.int = FALSE)
```




```
aids_fit_time <- survfit(Surv(time, censor) ~ sex, data=aids)
ggsurvplot(aids_fit_time, data=aids,  conf.int = TRUE)
```




```
aids_fit_time.d <- survfit(Surv(time_d, censor_d) ~ sex, data=aids)
ggsurvplot(aids_fit_time.d, data=aids,  conf.int = TRUE)
```



Survival Analysis

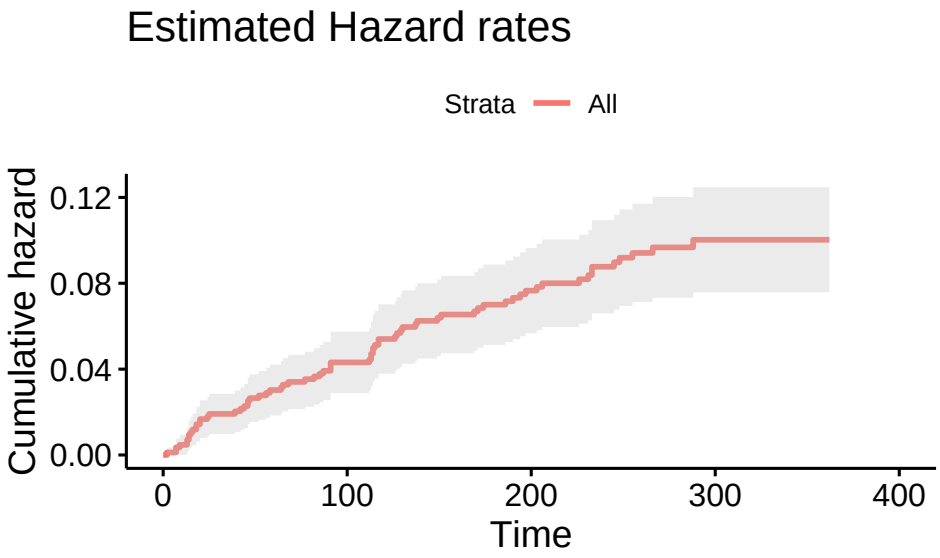
```
library(survival)
library (survminer)
library(ggplot2)
library(broom)

coxph(Surv(time_d,censor_d) ~ sex , data=aids)  %>% tidy()
```

```
## # A tibble: 1 x 7
##   term  estimate std.error statistic p.value conf.low conf.high
##   <chr>    <dbl>     <dbl>     <dbl>   <dbl>    <dbl>     <dbl>
## 1 sex      0.390     0.559     0.697   0.486   -0.706      1.49
```

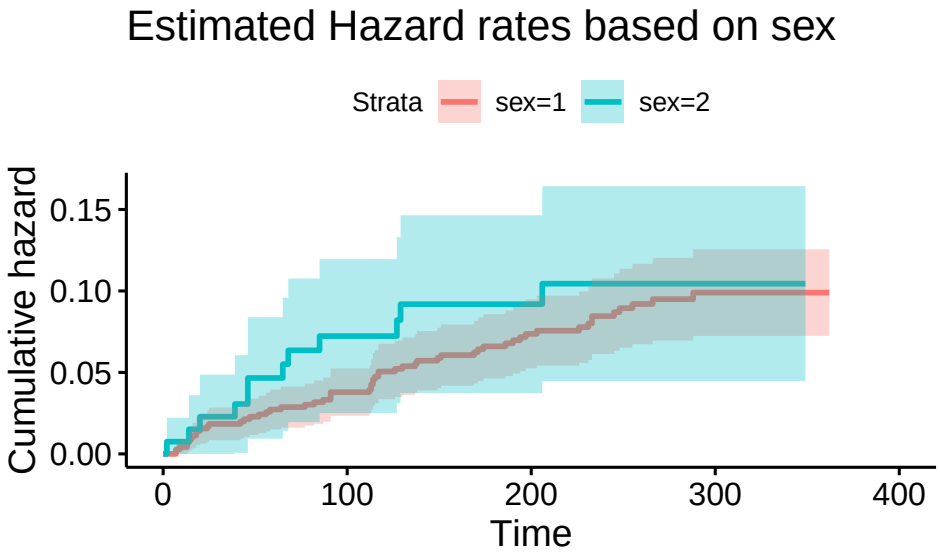
```r
coxph(Surv(time,censor) ~ sex, data=aids) %>% tidy()
```

```
## # A tibble: 1 x 7
##   term  estimate std.error statistic p.value conf.low conf.high
##   <chr>    <dbl>     <dbl>     <dbl>   <dbl>    <dbl>     <dbl>
## 1 sex      0.199     0.318     0.625   0.532   -0.424     0.821
```

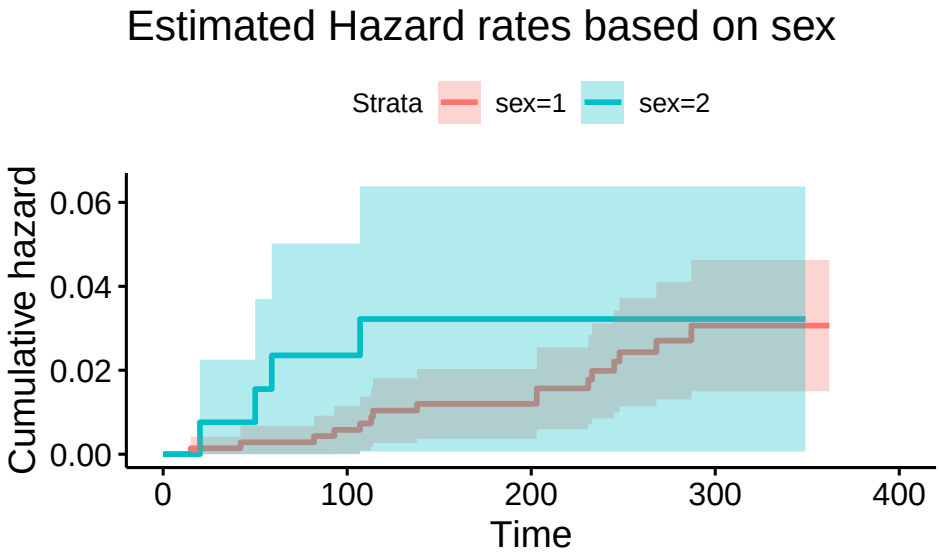
```r
ggsurvplot(survfit(Surv(time,censor) ~ 1, data=aids),
           censor=F, conf.int=T, fun="cumhaz") + ggtitle("Estimated Hazard rates")
```




```r
ggsurvplot(survfit(Surv(time,censor) ~ sex, data=aids),
           censor=F, conf.int=T, fun="cumhaz") + ggtitle("Estimated Hazard rates based on sex")
```




```r
ggsurvplot(survfit(Surv(time_d,censor_d) ~ sex, data=aids),
           censor=F, conf.int=T, fun="cumhaz") + ggtitle("Estimated Hazard rates based on sex")
```




```
#how tow modify so that sex is labeled as male and female
```

Juste's "Something New"