drug.r

April 26, 2019

In [48]:
```
library(ggplot2)
library(plotly)
d=read.csv("COAD_clinical_drug.csv",header=T)
```

In [49]:
```
table(d1[,"total_dose_units"])
#1 unknow , normalize 4th & 5th , essential for distance/gradient based algorithm
```

```
      g/day      mg mg/day  mg/kg  mg/m2      ug
   39     1     274      1      2      9      1
```

In [50]:
```
# count the number of cycles or doses the patient got of each medication.

d1=d[-which(d[,"number_cycles"]==""),]  # patietns with dosage &cycle info
d2=d1[-which(d1[,"total_dose"]==""),] # look for patietns with dosage info first
length(unique(d[,"bcr_patient_barcode"])) # assuming 1-1 mapping
length(unique(d1[,"bcr_patient_barcode"]))
length(unique(d2[,"bcr_patient_barcode"]))
```

```
154
69
58
```

In [51]:
```
#We can then also count the number of appropriate doses the patients got. Ie, did the
d[3,c("total_dose","total_dose_units","prescribed_dose","prescribed_dose_units","numb
# prescribed dose can be range or a single number
7185/12
```

	total_dose	total_dose_units	prescribed_dose	prescribed_dose_units	number_cycles
3	7185	mg	450-735	mg	12

```
598.75
```

In [60]:
```
#d2=d2[-which(d2["prescribed_dose"]==""),]

for(i in nrow(d2)){
    #print(d2[i,"total_dose_units"])
    #print(d2[i,"prescribed_dose_units"])
```

```
            temp=unlist(strsplit(x=as.character(d2[i,"prescribed_dose"]),split="-"))
            if(length(temp)==0){
                min_i=as.numeric(temp)*d2[,"number_cycles"]*0.9
                max_i=as.numeric(temp)*d2[,"number_cycles"]*1.1
            }
            else{
                 min_i=as.numeric(temp[0])
                 max_i=as.numeric(temp[1])

            }
            if(as.numeric(as.character(d2[i,"total_dose"]))<min_i || as.numeric(as.character(d
                warnings("not in range")
            }
        }
```

In [ ]: