Homework #5

Fan Yang, fy2232

December 8, 2017

Part I

1.

```
nodes <- read.csv("ckm_nodes.csv", header = T, as.is = T)

sum(na.omit(nodes$adoption_date) < Inf)
```

```
## [1] 109
```

```
sum(na.omit(nodes$adoption_date)==Inf)
```

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

```
length(nodes$adoption_date) - length(na.omit(nodes$adoption_date))
```

```
## [1] 121
```

2.

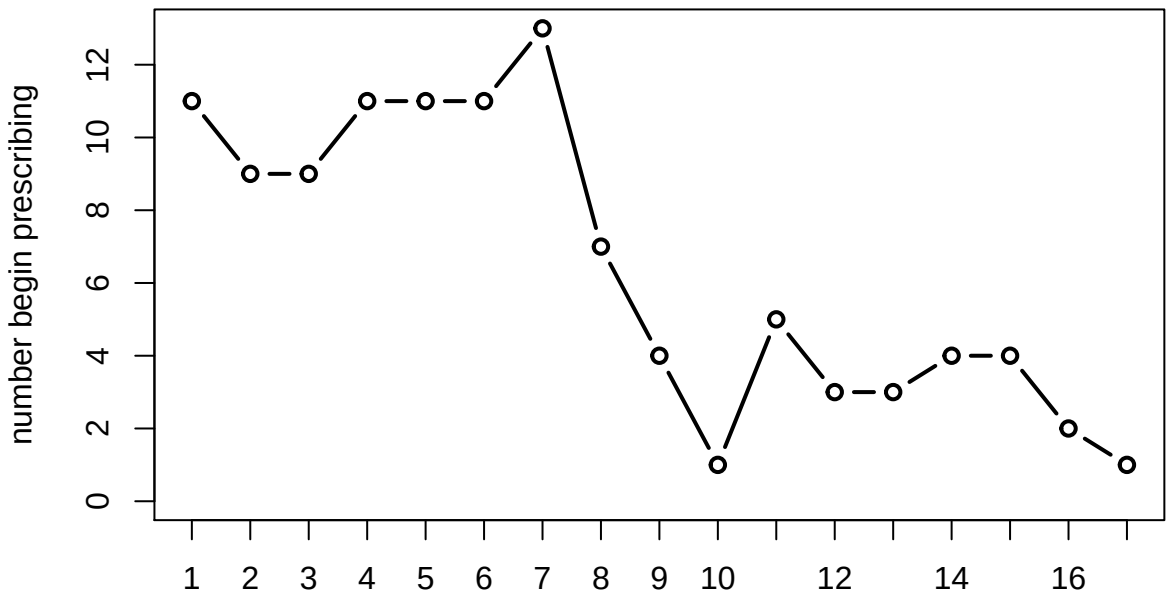
```
ind.notNA <- complete.cases(nodes$adoption_date)
ind.notNA <- seq(1:nrow(nodes))[ind.notNA]
length(ind.notNA)
```

```
## [1] 125
```

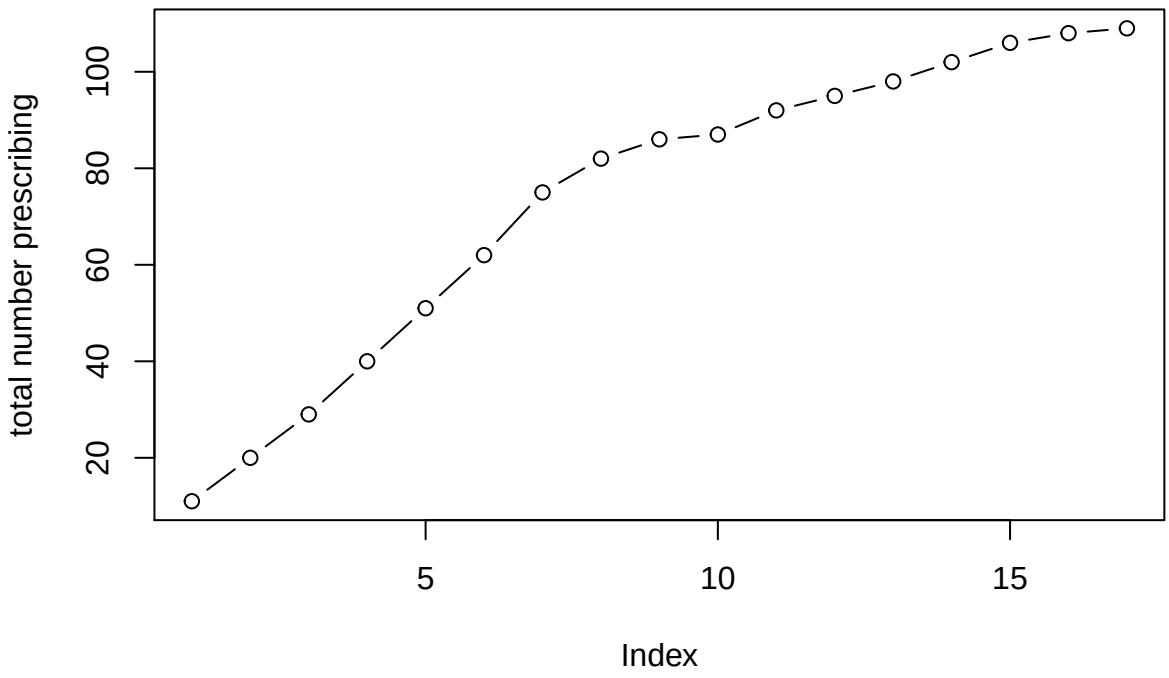
```
nodes <- nodes[ind.notNA,]
```

3.

```
num_mon <- table(nodes$adoption_date)[-18]
plot(num_mon,type = 'b',ylab="number begin prescribing")
```




```
plot(cumsum(num_mon),type = 'b',ylab="total number prescribing")
```



4.

```
begin2 <- nodes$adoption_date<=2
begin2 <- seq(1:nrow(nodes))[begin2]
begin2
```

```
##  [1]   1  10  13  20  27  45  48  55  56  63  66  70  71  73  74  75  76
## [18]  81  87 107
```

```
begin14 <- nodes$adoption_date>14
begin14 <- seq(1:nrow(nodes))[begin14]
begin14
```

```
##  [1]   7  14  16  17  30  39  42  50  52  62  67  79  82  85  88  89  91
## [18]  94  96  97 108 109 125
```

5.

```
adopters <- function(month, not.yet = FALSE){
  index <- seq(1:nrow(nodes))
  late <- index[nodes$adoption_date>month]
  early <- index[nodes$adoption_date==month]
  if(not.yet){
    return(late)
  }else{
```

```
    return(early)
  }
}
adopters(2)
```

```
## [1]  10  13  20  56  71  75  76  87 107
```

```
length(adopters(2))
```

```
## [1] 9
```

```
length(adopters(month = 14, not.yet = TRUE))
```

```
## [1] 23
```

Part II

6.

```
network <- read.table("ckm_network.txt")
dim(network)
```

```
## [1] 246 246
```

```
network <- network[ind.notNA,ind.notNA]
names(network) <- c()
dim(network)
```

```
## [1] 125 125
```

7.

```
doc.contacts <- apply(network,2,sum)
doc.contacts[41]
```

```
## [1] 3
```

8.

```
friend37 <- (network[,37]==1)*(nodes$adoption_date<=5)
sum(friend37)
```

```
## [1] 3
```

```
sum(friend37) / doc.contacts[37]
```

```
## [1] 0.6
```

9.

```
count_peer_pressure <- function(index, month){
  return (sum((network[,index]==1)*(nodes$adoption_date<=month)))
}
count_peer_pressure(37,5)
```

```
## [1] 3
```

10.

```
# approach 1
# prop_peer_pressure <- function(index, month){
#   return (
#     sum((network[,index]==1)*(nodes$adoption_date<=month)) /
#     sum((network[,index]==1))
#   )
# }
# prop_peer_pressure(37,5)
# prop_peer_pressure(102,12)

# approach 2
prop_peer_pressure <- function(index, month){
  return (count_peer_pressure(index,month) / doc.contacts[index])
}
prop_peer_pressure(37,5)
```

```
## [1] 0.6
```

```
prop_peer_pressure(102,12)
```

```
## [1] NaN
```

11.

```
average_contact <- function(month){
  return( c(
    mean(sapply(adopters(month,FALSE),prop_peer_pressure,month=month),na.rm=TRUE),
    mean(sapply(adopters(month,TRUE),prop_peer_pressure,month=month),na.rm=TRUE)  )
  )
}
average_contact(2)
```

```
## [1] 0.2370851 0.1584696
```

12.

```
month.ind <- 1:17
begin_ave <- sapply(month.ind,average_contact)[1,]
later_ave <- sapply(month.ind,average_contact)[2,]

library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.2
```

```
ggplot() +
  geom_line(aes(x =month.ind , y = begin_ave,col='red')) +
  geom_line(aes(x =month.ind , y = later_ave,col='blue')) +
```

```
scale_colour_discrete(labels = c('later','begin')) +
xlab("month") + ylab("average proportion")
```