

HW9a

Autumn Li

12/4/2016

Part I

1

- (a) How many doctors began prescribing tetracycline in each month of the study? (b) How many never prescribed it? (c) How many are NAs?

```
ckm_nodes.1 <- read.csv("~/Desktop/ckm_nodes-1.csv")
ckm= ckm_nodes.1
date = ckm$adoption_date
na = length(date[date == "NA"]);na

## [1] 121

never = length(na.omit(date[date == "Inf"]));never

## [1] 16

begin = length(date)- length(date[date == "NA"]) - length(na.omit(date[date == "Inf"]));begin

## [1] 109
```

- (a) There are 109 doctors begin prescribing tetracycline in each month of the study; (b) There are 16 doctors never prescribing it. (c) There are 121 NA.

2. Create a vector which records the index numbers of the doctors for whom adoption date is not NA.

```
not_na = na.omit(date)
ind_na = which(is.element(date=="NA",date))
vect = 1:246
dr_ind = vect[-ind_na]
```

3. Create a plot of the number of doctors who began prescribing tetracycline each month versus time.

```
ind_inf = which(is.element(not_na=="Inf",not_na))
dr_num = not_na[-ind_inf]
dr_num = sort(dr_num);dr_num

## [1] 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 3 3 3
## [24] 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5
## [47] 5 5 5 5 5 6 6 6 6 6 6 6 6 6 6 6 7 7 7 7 7 7
## [70] 7 7 7 7 7 8 8 8 8 8 8 8 8 9 9 9 9 10 11 11 11 11
## [93] 12 12 12 13 13 13 14 14 14 14 15 15 15 15 16 16 17
```

```

count2 = NULL
i=1
while(i <= max(dr_num)){
  count1 = NULL
  for(k in 1:length(dr_num)){
    if(dr_num[k] == i){
      count1[k] = 0
    }
  }
  count2[i] = length(na.omit(count1))
  i = i +1
}
count2

```

```
## [1] 11 9 9 11 11 11 13 7 4 1 5 3 3 4 4 2 1
```

```

ind_inf = which(is.element(not.na=="Inf",not.na))
dr_num = not.na[-ind_inf]
dr_num = sort(dr_num);dr_num

```

```

## [1] 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 3 3 3
## [24] 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5
## [47] 5 5 5 5 5 6 6 6 6 6 6 6 6 6 6 6 7 7 7 7 7 7
## [70] 7 7 7 7 7 7 8 8 8 8 8 8 8 8 9 9 9 10 11 11 11 11
## [93] 12 12 12 13 13 13 14 14 14 14 15 15 15 15 16 16 17

```

```

count3 = NULL
count4 = NULL
i=1
while(i <= max(dr_num)){
  for(k in 1:length(dr_num)){
    if(dr_num[k] == i){
      count4[k] = 0
    }
  }
  count3[i] = length(na.omit(count4))
  i = i +1
}
count3

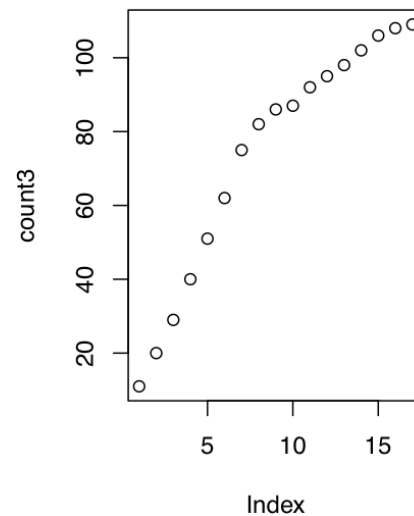
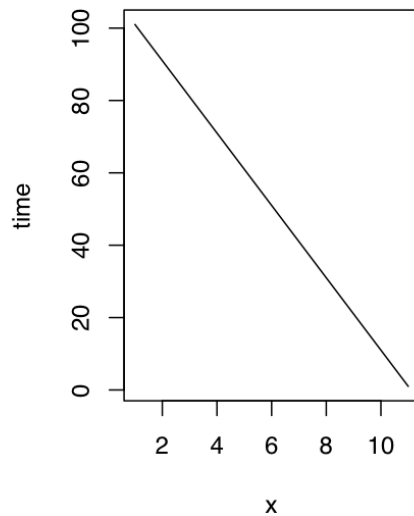
```

```
## [1] 11 20 29 40 51 62 75 82 86 87 92 95 98 102 106 108 109
```

```

par(mfrow=c(1,2))
plot(time,count2)
plot(count3)

```



4.Cre-

ate a logical vector which indicates for each doctor whether they had begun pre- scribing tetracycline by month 2.

```
logic = NULL
for(i in 1:length(not.na)){
  if(not.na[i]<=2) {
    logic[i] = 0
  }
}
vect1 = 1:length(logic)
ind_m2 = vect1[-which(is.element(logic == 0,logic))]
length(ind_m2)
```

```
## [1] 20
```

```
logic2 = NULL
for(i in 1:length(not.na)){
  if(not.na[i]>14) {
    logic2[i] = 0
  }
}
vect2 = 1:length(logic2)
ind_m14 = vect2[-which(is.element(logic2 == 0,logic2))]
length(ind_m14)
```

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

Part II

5.

```
ckm_network.1 <- read.table("~/Desktop/ckm_network-1.txt", quote="\")
ckm_txt = ckm_network.1
dim(ckm_txt)
```

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

```
network = ckm_txt[-ind_na,-ind_na]
dim(network)
```

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

6. Create a vector which stores the number of contacts each doctor has. Do not use a loop. Check that doctor number 41 has 3 contacts.

```
apply(network,1,sum)[41]
```

```
## 70
## 3
```

7.

```
nodes = read.csv("~/Desktop/ckm_nodes-1.csv")
date = nodes$adoption_date
not_na = na.omit(date)
ind_na = which(is.element(date=="NA",date))
vect = 1:246
dr_ind = vect[-ind_na]
nodes = nodes[dr_ind,]
doctor_37 = network[,37]
m = sum(nodes$adoption_date <= 5 * network[,37])
propotion = m/sum(network[,37])
propotion
```

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