Lab06

Yiheng

7/17/2020

How the Tetracycline Came to Peoria

`ckm_nodes.csv` has information about each individual doctor in the four towns. `ckm_network.dat` records which doctors knew each other.

Part I

1.

```
ckm_nodes <- read_csv('data/ckm_nodes.csv')
```

```
## Parsed with column specification:
## cols(
##   city = col_character(),
##   adoption_date = col_double(),
##   medical_school = col_character(),
##   attend_meetings = col_character(),
##   medical_journals = col_double(),
##   free_time_with = col_character(),
##   discuss_medicine_socially = col_character(),
##   club_with_drs = col_character(),
##   drs_among_three_best_friends = col_double(),
##   practicing_here = col_character(),
##   office_visits_per_week = col_character(),
##   proximity_to_other_drs = col_character(),
##   specialty = col_character()
## )
```

2.

a.

```
table(ckm_nodes$adoption_date)
```

```
##
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17 Inf
##  11   9   9  11  11  11  13   7   4   1   5   3   3   4   4   2   1  16
```

```
sum(is.na(ckm_nodes$adoption_date))
```

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

⇒ 16 doctors never prescribed, 121 are NAs.

b.

```
anum <- which(!is.na(ckm_nodes$adoption_date))
length(anum)
```

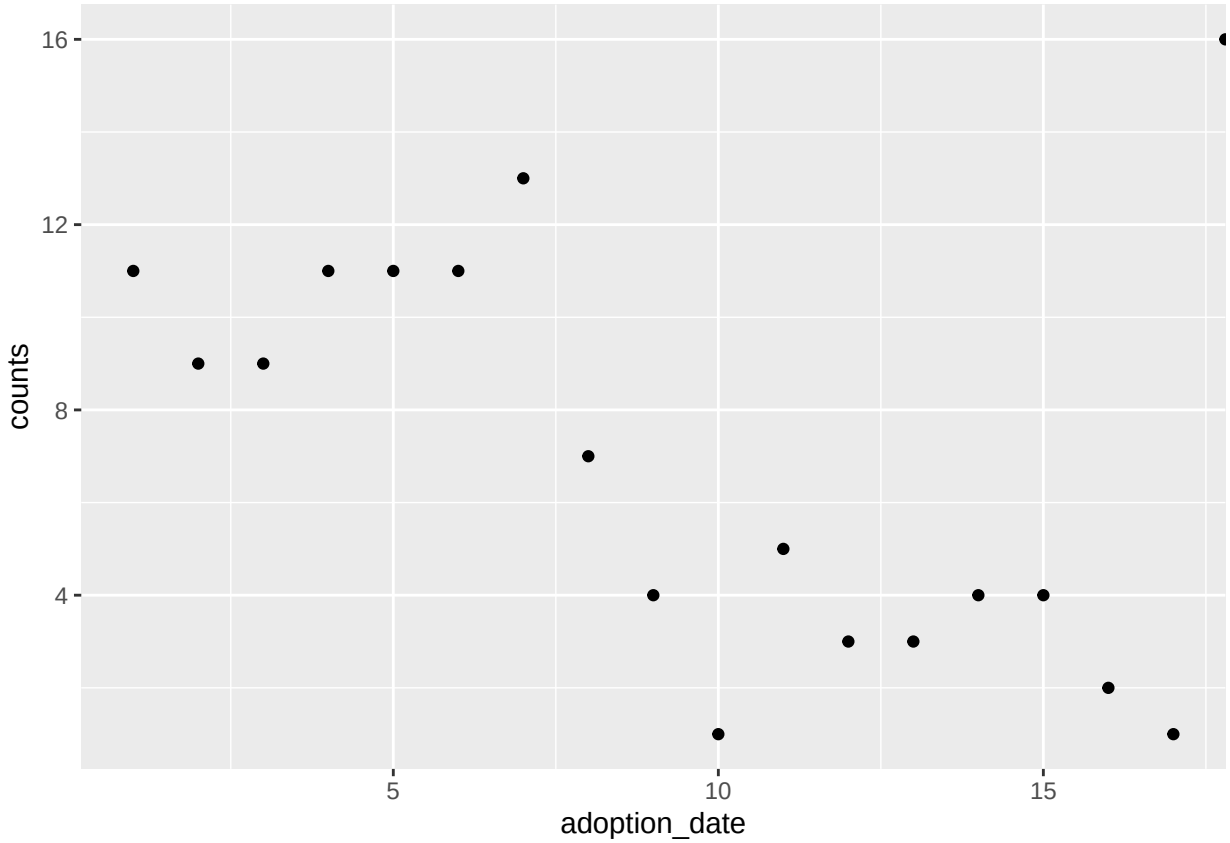
```
## [1] 125
```

```
ckm_nodes <- ckm_nodes %>%
  filter(!is.na(adoption_date))
```

3.

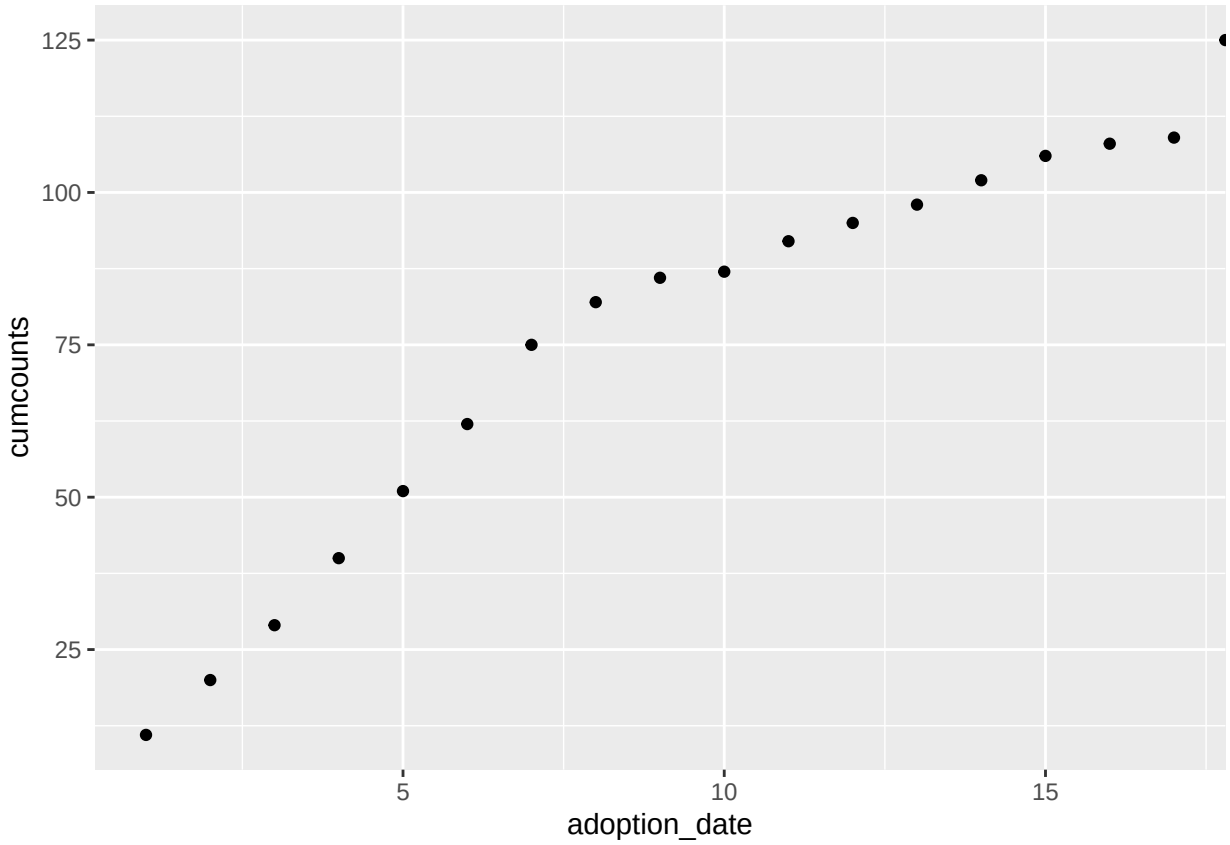
```
ckm_nodes %>% group_by(adoption_date) %>%
  summarize(counts = n()) %>%
  ggplot() + #pipe operator exists, shouldn't write ggplot(data = ckm_nodes)
  geom_point(mapping = aes(x = adoption_date, y = counts))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



```
ckm_nodes %>% group_by(adoption_date) %>%
  summarize(counts = n()) %>%
  arrange(adoption_date) %>%
  mutate(cumcounts = cumsum(counts)) %>%
  ggplot() +
  geom_point(aes(x = adoption_date, y = cumcounts))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



4.a.

```
adopted.early.boo <- ckm_nodes$adoption_date <= 2
sum(adopted.early.boo)
```

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

```
adopted.early.ind <- which(ckm_nodes$adoption_date <= 2)
```

b.

```
adopted.late.boo <- ckm_nodes$adoption_date > 14
sum(adopted.late.boo)
```

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

```r
adopted.late.ind <- which(ckm_nodes$adoption_date >14)
```

Part II

5.

```r
ckm_network <- read.table('data/ckm_network.dat')
#read_table() gives warning messages but read.table() doesn't (?)
dim(ckm_network)
```

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

```r
ckm_network <- ckm_network[anum,anum]
dim(ckm_network)
```

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

6.

```r
contact.num <- rowSums(ckm_network)
contact.num[41]
```

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

7.

a.

```r
contact.37 <- ckm_network[37,]==1 & ckm_nodes$adoption_date<=5
sum(contact.37)
```

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

b.

```r
sum(contact.37)/sum(ckm_network[37,])
```

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