STAT346_HW3_2019150445

2019150445/Shin Baek Rok

2020 10 30

1.

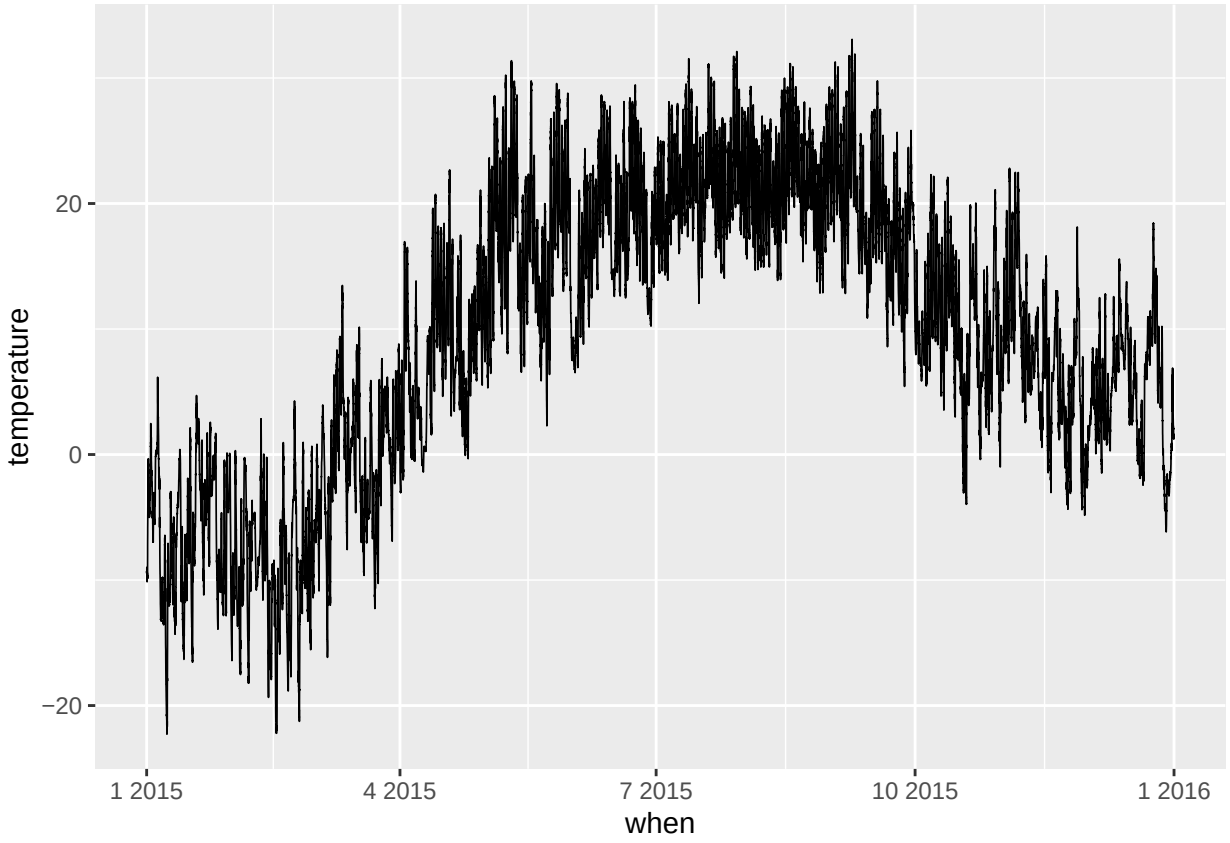
```
library(macleish)
```

```
## Warning: package 'macleish' was built under R version 4.0.3
```

```
## Loading required package: etl
```

```
## Warning: package 'etl' was built under R version 4.0.3
```

```
whately_2015 %>% ggplot()+geom_line(aes(x=when,y=temperature),size=0.3)
```



2.a)

```
library(nasaweather)
```

```
## Warning: package 'nasaweather' was built under R version 4.0.3
```

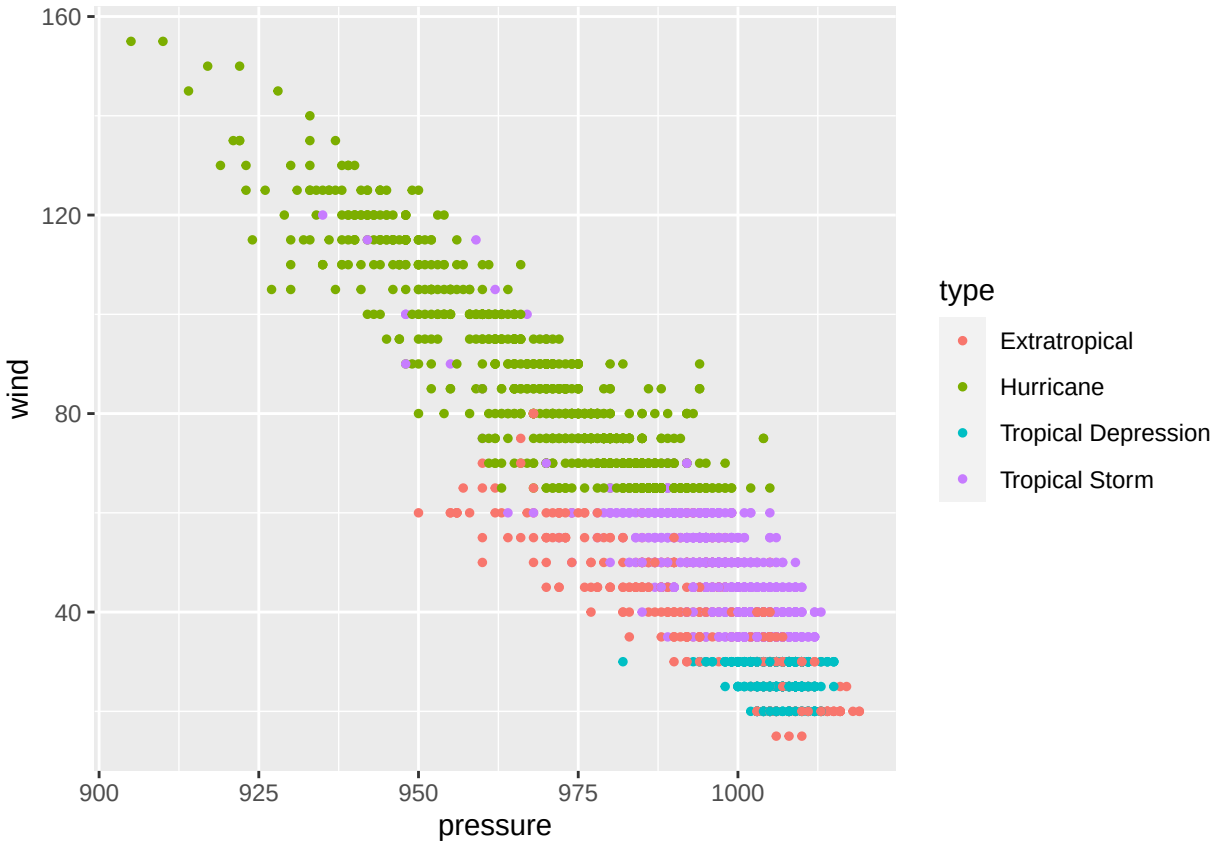
```
##
## Attaching package: 'nasaweather'
```

```
## The following object is masked from 'package:dplyr':
##
##     storms
```

```
storms %>% glimpse()
```

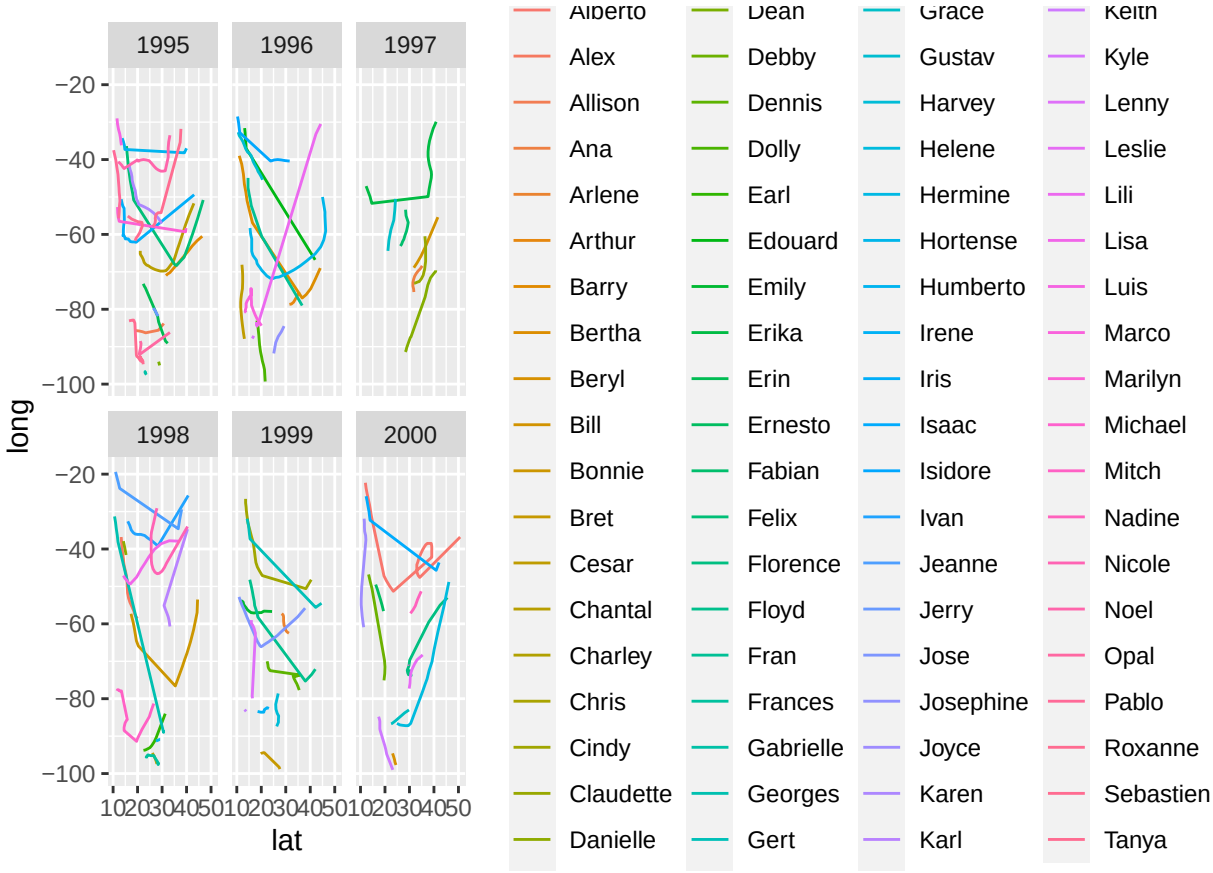
```
## Rows: 2,747
## Columns: 11
## $ name     <chr> "Allison", "Allison", "Allison", "Allison", "Allison", "Al...
## $ year     <int> 1995, 1995, 1995, 1995, 1995, 1995, 1995, 1995, 1995, 1995...
## $ month    <int> 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6...
## $ day      <int> 3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 7...
## $ hour     <int> 0, 6, 12, 18, 0, 6, 12, 18, 0, 6, 12, 18, 0, 6, 12, 18, 0,...
## $ lat      <dbl> 17.4, 18.3, 19.3, 20.6, 22.0, 23.3, 24.7, 26.2, 27.6, 28.5...
## $ long     <dbl> -84.3, -84.9, -85.7, -85.8, -86.0, -86.3, -86.2, -86.2, -8...
## $ pressure <int> 1005, 1004, 1003, 1001, 997, 995, 987, 988, 988, 990, 990,...
## $ wind     <int> 30, 30, 35, 40, 50, 60, 65, 65, 65, 60, 60, 45, 30, 35, 35...
## $ type     <chr> "Tropical Depression", "Tropical Depression", "Tropical St...
## $ seasday  <int> 3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 7...
```

```
storms %>% ggplot(aes(x=pressure,y=wind,col=type))+
  geom_point(size=1)
```

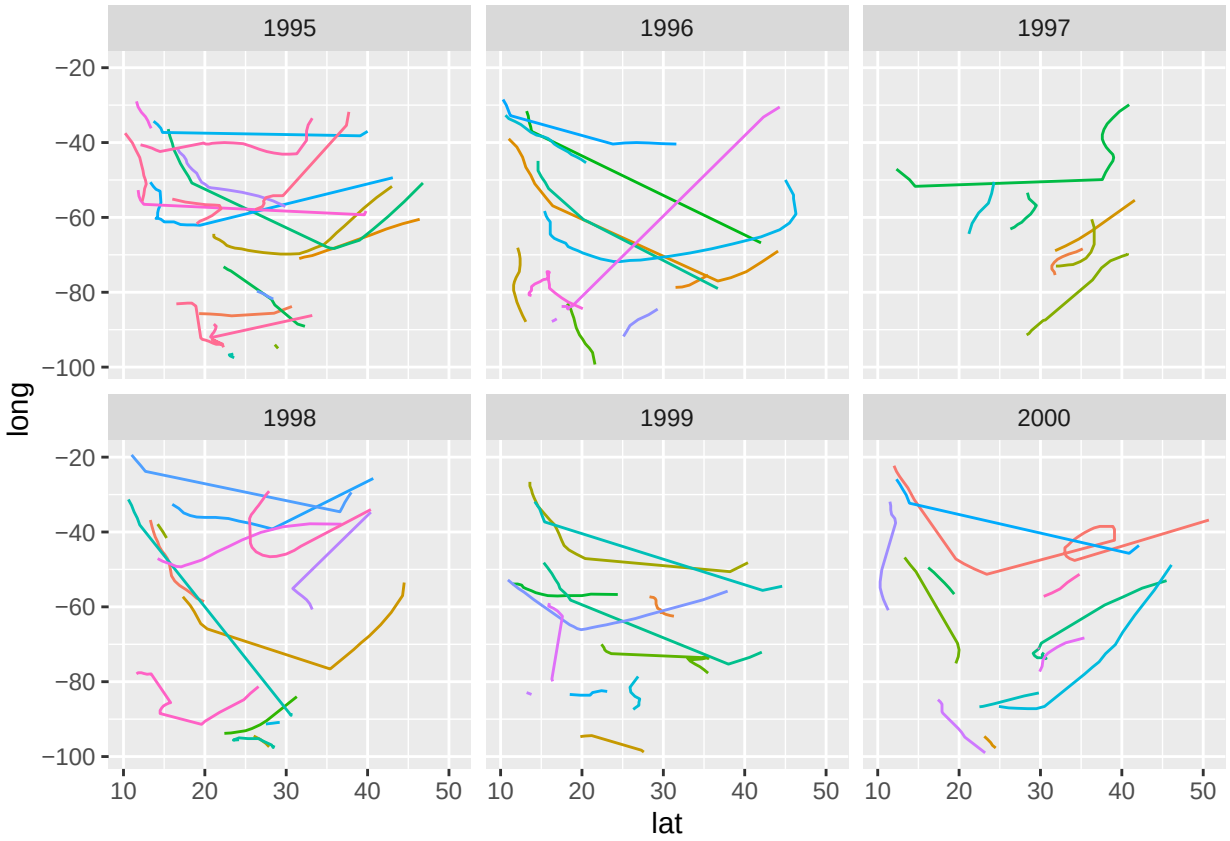


2.b)

```
storms %>% filter(type=='Tropical Storm') %>% ggplot(aes(x=lat,y=long,col=name,label=name))+geom_path()+
  facet_wrap(~year)
```

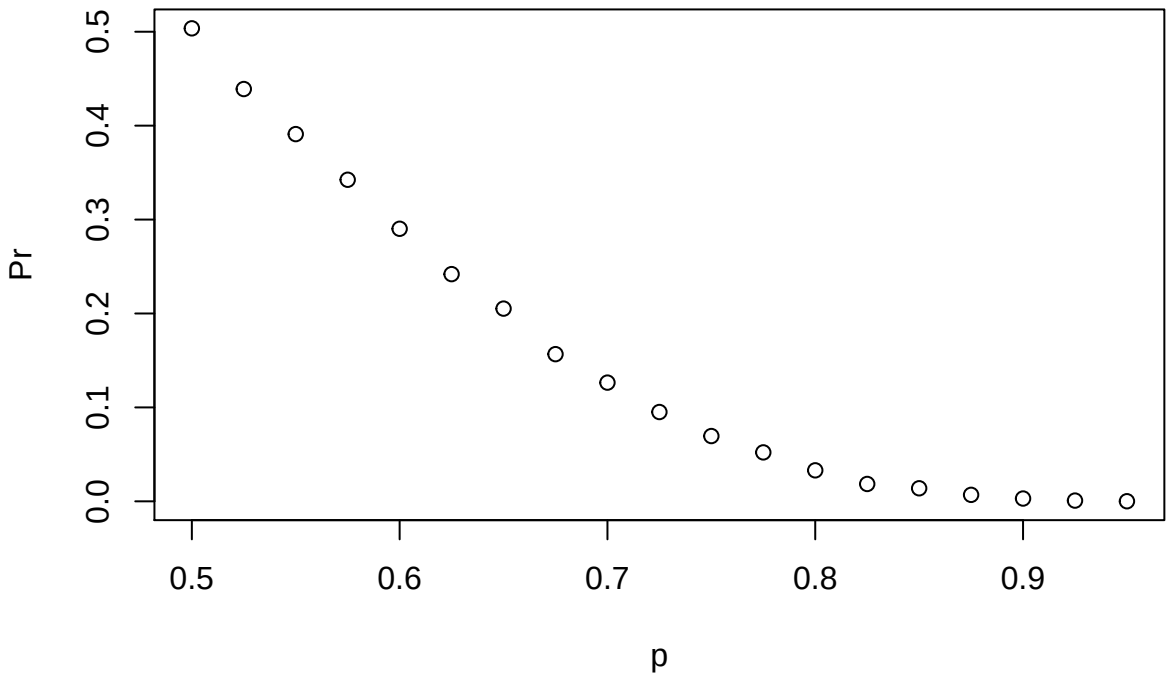



```
#legend=FALSE
storms %>% filter(type=='Tropical Storm') %>% ggplot(aes(x=lat,y=long,col=name,label=name))+geom_path()+
  facet_wrap(~year)+theme(legend.position='none')
```



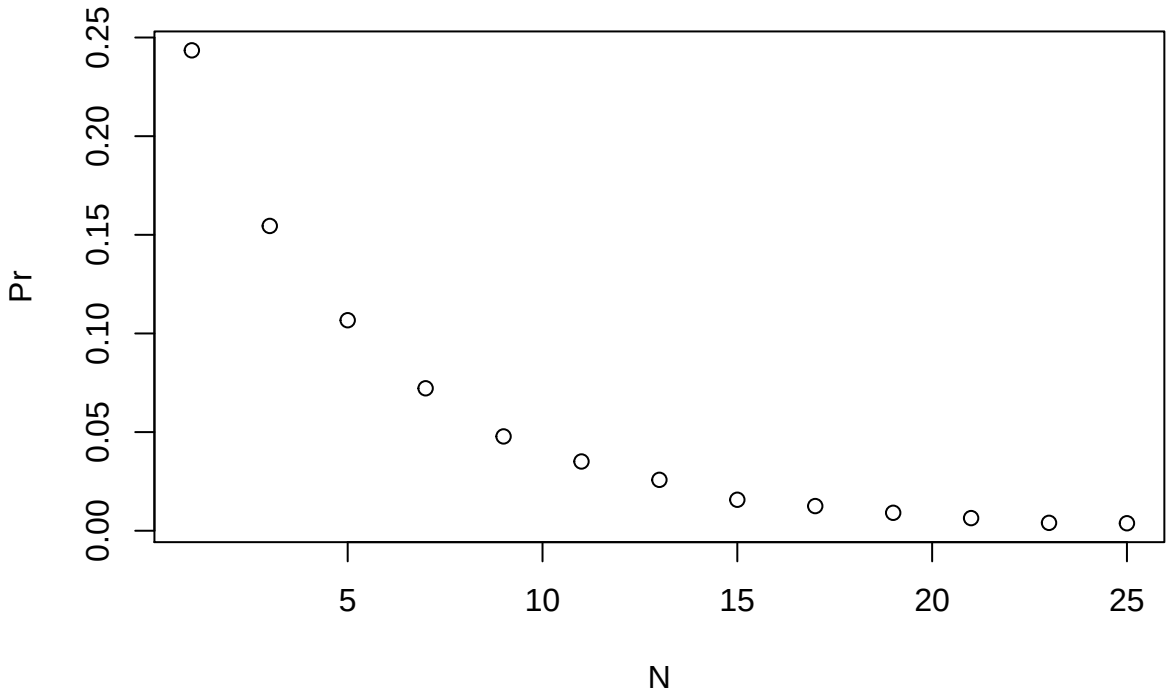
3.a)

```
prob_win<-function(p){
  B<-10000
  result<-replicate(B,{
    b_win<-sample(c(1,0),7,replace=TRUE, prob=c(1-p,p))
    sum(b_win)>=4
  })
  mean(result)
}
p<-seq(0.5,0.95,0.025)
Pr<-sapply(p,prob_win)
plot(p,Pr)
```



3.b)

```
prob_win<-function(N,p=0.75){
  B<-10000
  result<-replicate(B,{
    b_win<-sample(c(1,0),N,replace=TRUE,prob=c(1-p,p))
    sum(b_win)>=(N+1)/2
  })
  mean(result)
}
N<-seq(1,25,2)
Pr<-sapply(N,prob_win)
plot(N,Pr)
```



4.1.

```
library(dslabs)
data('polls_us_election_2016')
polls<-polls_us_election_2016 %>% filter(enddate >= '2016-10-31' & state=='U.S.')
N<-polls$samplesize[1]
x_hat<-polls$rawpoll_clinton[1]/100
se_hat<-sqrt(x_hat*(1-x_hat)/N)
CI_95<-c(x_hat-qnorm(0.975)*se_hat,x_hat+qnorm(0.975)*se_hat)
CI_95
```

```
## [1] 0.4492385 0.4907615
```

4.2.

```
polls<-polls %>% mutate(x_hat=rawpoll_clinton/100) %>%
  mutate(se_hat=sqrt(x_hat*(1-x_hat)/samplesize)) %>%
  mutate(lower=x_hat-se_hat*qnorm(0.975),
  upper=x_hat+se_hat*qnorm(0.975)) %>% select(pollster, enddate, x_hat, lower, upper)
head(polls)
```

```
##                                                    pollster    enddate  x_hat
## 1                                ABC News/Washington Post 2016-11-06 0.4700
## 2                                 Google Consumer Surveys 2016-11-07 0.3803
## 3                                                   Ipsos 2016-11-06 0.4200
## 4                                                  YouGov 2016-11-07 0.4500
## 5                                          Gravis Marketing 2016-11-06 0.4700
## 6 Fox News/Anderson Robbins Research/Shaw & Company Research 2016-11-06 0.4800
##       lower     upper
## 1 0.4492385 0.4907615
## 2 0.3744632 0.3861368
## 3 0.3993524 0.4406476
## 4 0.4339199 0.4660801
## 5 0.4624165 0.4775835
## 6 0.4527896 0.5072104
```

4.3, 4.4

```
polls<-polls %>% mutate(hit=(lower<0.482 & upper>0.482))
mean(polls$hit)
```

```
## [1] 0.3142857
```

4.5

95%

4.6

```
polls <- polls_us_election_2016 %>%
  filter(enddate >= "2016-10-31" & state == "U.S.")  %>%
  mutate(d_hat = rawpoll_clinton / 100 - rawpoll_trump / 100)
N<-polls$samplesize[1]
d_hat<-polls$d_hat[1]
se_hat<-2*sqrt(((d_hat+1)/2)*(1-(d_hat+1)/2)/N)#d=2p-1
CI95<-c(d_hat-qnorm(0.975)*se_hat,d_hat+qnorm(0.975)*se_hat)
CI95
```

```
## [1] -0.001564627  0.081564627
```

5.1

```
library(HistData)
```

```
## Warning: package 'HistData' was built under R version 4.0.3
```

```
library(broom)
```

```
## Warning: package 'broom' was built under R version 4.0.3
```

```
data('GaltonFamilies')
set.seed(1)
galton_heights <- GaltonFamilies %>%
  group_by(family, gender) %>%
  sample_n(1) %>%
  ungroup()

galton_heights %>%
  gather(parent, parentHeight, father:mother) %>%
  mutate(child = ifelse(gender == "female", "daughter", "son")) %>%
  unite(pair, c("parent", "child")) %>% group_by(pair) %>%
  do(tidy(lm(childHeight~parentHeight,data=.)))%>%
  filter(term=='parentHeight')
```

```
## # A tibble: 4 x 6
## # Groups:   pair [4]
##   pair            term         estimate std.error statistic  p.value
##   <chr>           <chr>           <dbl>     <dbl>     <dbl>    <dbl>
## 1 father_daughter parentHeight    0.345    0.0625      5.52 1.21e- 7
## 2 father_son      parentHeight    0.426    0.0646      6.59 4.74e-10
## 3 mother_daughter parentHeight    0.413    0.0745      5.55 1.07e- 7
## 4 mother_son      parentHeight    0.312    0.0739      4.22 3.84e- 5
```

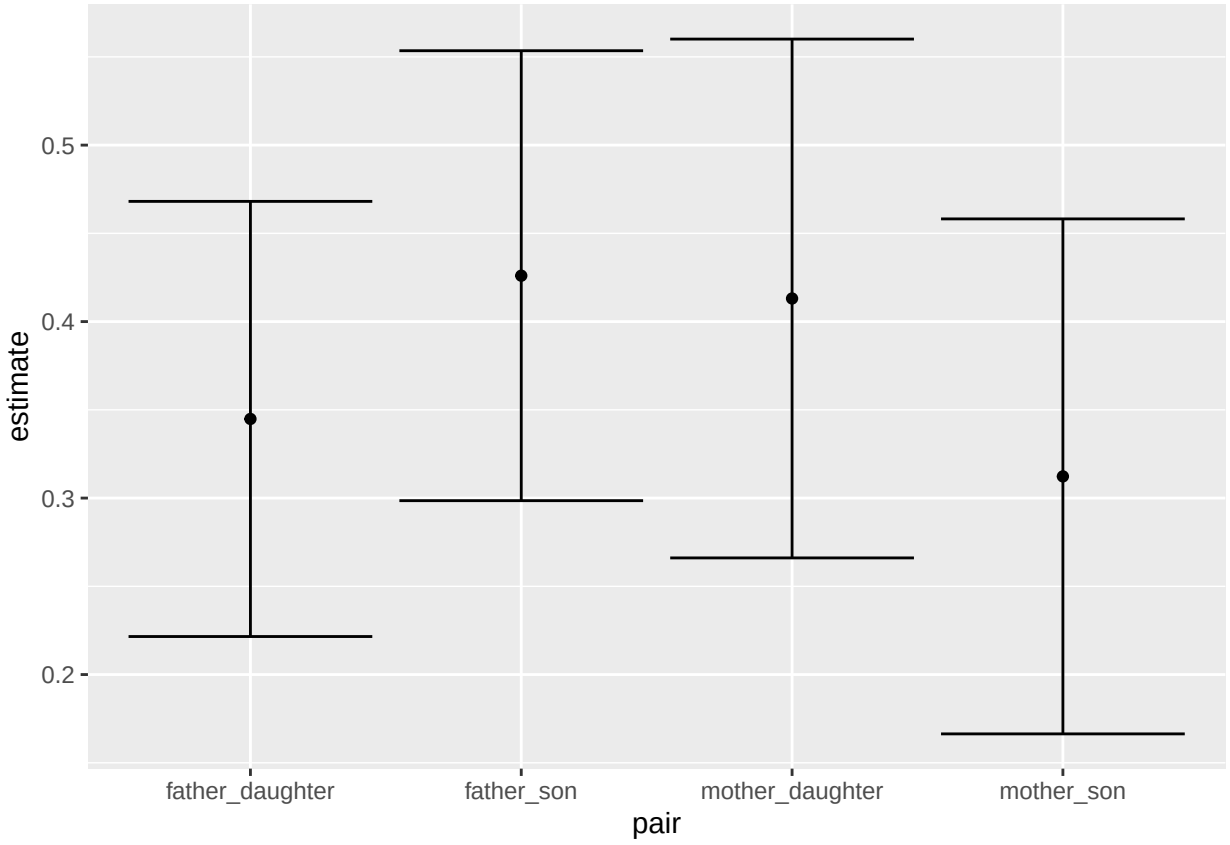
5.2

```
galton_heights %>%
  gather(parent, parentHeight, father:mother) %>%
  mutate(child = ifelse(gender == "female", "daughter", "son")) %>%
  unite(pair, c("parent", "child")) %>% group_by(pair) %>%
  do(tidy(lm(childHeight~parentHeight,data=.),conf.int=TRUE))%>%
  filter(term=='parentHeight')
```

```
## # A tibble: 4 x 8
## # Groups:   pair [4]
##   pair        term      estimate std.error statistic  p.value conf.low conf.high
##   <chr>       <chr>        <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
## 1 father_dau~ parentHe~    0.345    0.0625      5.52 1.21e- 7    0.222     0.468
## 2 father_son  parentHe~    0.426    0.0646      6.59 4.74e-10    0.299     0.553
## 3 mother_dau~ parentHe~    0.413    0.0745      5.55 1.07e- 7    0.266     0.560
## 4 mother_son  parentHe~    0.312    0.0739      4.22 3.84e- 5    0.166     0.458
```

5.3

```
galton_heights %>%
  gather(parent, parentHeight, father:mother) %>%
  mutate(child = ifelse(gender == "female", "daughter", "son")) %>%
  unite(pair, c("parent", "child")) %>% group_by(pair) %>%
  do(tidy(lm(childHeight~parentHeight,data=.),conf.int=TRUE))%>%
  filter(term=='parentHeight') %>%
  ggplot(aes(x=pair,y=estimate,ymin=conf.low,ymax=conf.high))+geom_errorbar()+geom_point()
```



6.1

```
men=c(sum(research_funding_rates$awards_men),sum(research_funding_rates$applications_men)-sum(research_
women=c(sum(research_funding_rates$awards_women),sum(research_funding_rates$applications_women)-sum(res
table<-data.frame(men,women,row.names=c('awards','Not awards'))
table
```

```
##             men women
## awards      290   177
## Not awards 1345  1011
```

6.2

```
290/(290+1345)-177/(177+1011)
```

```
## [1] 0.02838013
```

6.3

```
chisq.test(table)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table
## X-squared = 3.8111, df = 1, p-value = 0.05091
```

6.4

```
research_funding_rates$discipline<-reorder(research_funding_rates$discipline,research_funding_rates$suc
research1<-research_funding_rates


research2<-gather(research1,var, value, 2:10)

research2[55:63,2]<-'successrates_total'
research2[64:72,2]<-'successrates_men'
research2[73:81,2]<-'successrates_women'

research_fin<-separate(research2,var,c('type','gender')) %>% spread(gender, value)
research_fin
```

```
##             discipline         type   men total women
## 1      Social sciences applications 425.0 834.0 409.0
## 2      Social sciences       awards  65.0 112.0  47.0
## 3      Social sciences successrates  15.3  13.4  11.5
## 4     Medical sciences applications 245.0 505.0 260.0
## 5     Medical sciences       awards  46.0  75.0  29.0
## 6     Medical sciences successrates  18.8  14.9  11.2
## 7    Interdisciplinary applications 105.0 183.0  78.0
## 8    Interdisciplinary       awards  12.0  29.0  17.0
## 9    Interdisciplinary successrates  11.4  15.8  21.8
## 10          Humanities applications 230.0 396.0 166.0
## 11          Humanities       awards  33.0  65.0  32.0
## 12          Humanities successrates  14.3  16.4  19.3
## 13  Technical sciences applications 189.0 251.0  62.0
## 14  Technical sciences       awards  30.0  43.0  13.0
```

```
## 15  Technical sciences successrates  15.9  17.1  21.0
## 16 Earth/life sciences applications 156.0 282.0 126.0
## 17 Earth/life sciences       awards  38.0  56.0  18.0
## 18 Earth/life sciences successrates  24.4  19.9  14.3
## 19   Physical sciences applications 135.0 174.0  39.0
## 20   Physical sciences       awards  26.0  35.0   9.0
## 21   Physical sciences successrates  19.3  20.1  23.1
## 22   Chemical sciences applications  83.0 122.0  39.0
## 23   Chemical sciences       awards  22.0  32.0  10.0
## 24   Chemical sciences successrates  26.5  26.2  25.6
## 25            Physics applications  67.0  76.0   9.0
## 26            Physics       awards  18.0  20.0   2.0
## 27            Physics successrates  26.9  26.3  22.2
```

6.5

```
research_fin[c(3,6,9,12,15,18,21,24,27),] %>% ggplot() +geom_point(aes(x=discipline,y=men),col='red')+g
```