Homework 5

201982188 통계학과 박현주

2020-10-20

Your turn 1.

For this your turn use the french_fries data from the reshape2 package:
data("french_fries", package="reshape2")

1-1.

Use gather or pivot_longer from the tidyr package to combine the different
scales for assessing french fries into a single variable. Call the key-value
pair "scale" and "score".

```r
library(tidyverse)
data("french_fries", package="reshape2")

french_fries %>%
  pivot_longer(cols = potato:painty,
               names_to = "scale",
               values_to = "score") -> prob1

knitr::kable(prob1 %>% head(10))
```

time	treatment	subject	rep scale	score
1	1	3	1 potato	2.9
1	1	3	1 buttery	0.0
1	1	3	1 grassy	0.0
1	1	3	1 rancid	0.0
1	1	3	1 painty	5.5
1	1	3	2 potato	14.0
1	1	3	2 buttery	0.0
1	1	3	2 grassy	0.0
1	1	3	2 rancid	1.1
1	1	3	2 painty	0.0

1-2.

Use spread or pivot_wider from the tidyr package to get a format in which
you can directly compare values from week 1 to week 10.

```r
prob1 %>% pivot_wider(names_from = time, values_from = score) -> prob2

knitr::kable(prob2 %>% head(10))
```

treatment	subject	rep scale	1	2	3	4	5	6	7	8	9	10
1	3	1 potato	2.9	9.0	11.8	13.6	14.0	0.4	2.9	3.5	1.1	NA
1	3	1 buttery	0.0	0.3	0.2	0.1	0.3	1.2	0.0	0.5	0.4	NA

treatment	subject	rep	scale	1	2	3	4	5	6	7	8	9	10
1	3	1	grassy	0.0	0.1	0.0	0.0	0.0	0.0	0.0	1.3	0.0	NA
1	3	1	rancid	0.0	5.8	6.0	1.7	0.0	0.0	0.0	0.0	0.0	NA
1	3	1	painty	5.5	0.3	0.0	0.0	1.7	9.5	5.5	3.8	7.0	NA
1	3	2	potato	14.0	5.5	7.8	5.3	12.9	3.3	0.8	0.6	2.5	NA
1	3	2	buttery	0.0	0.5	0.5	0.0	0.8	1.1	0.0	0.3	0.5	NA
1	3	2	grassy	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NA
1	3	2	rancid	1.1	8.6	11.0	0.9	2.8	0.0	0.0	0.0	0.0	NA
1	3	2	painty	0.0	0.0	0.0	0.0	0.0	3.0	8.2	8.1	3.4	NA

1-3.

Plot a scatter plot of values in week 1 against week 10. Facet by treatment and scale, color by individuals and use different shapes for the replicates. Is there a pattern visible?

```
prob2 %>% mutate(row = as.factor(rep(1:24,5*3))) %>%
                ggplot(aes(x=`1`, y=`10`))+
  geom_point(aes(color=row,shape=as.factor(rep)))+
  geom_smooth(se=FALSE,size=0.5,color='green')+
  facet_grid(treatment~scale)+
  labs(title = "Scatter plot : Week 1 vs Week 10",shape='rep')+ #title and leg
  guides(color = FALSE)+
  xlab('Week 1')+ylab('Week 10')+
  theme(plot.title = element_text(hjust = 0.5,size = 15, face = "bold",margin=
        legend.title=element_text(size=12,hjust = 0.5)
        )#edit text
```
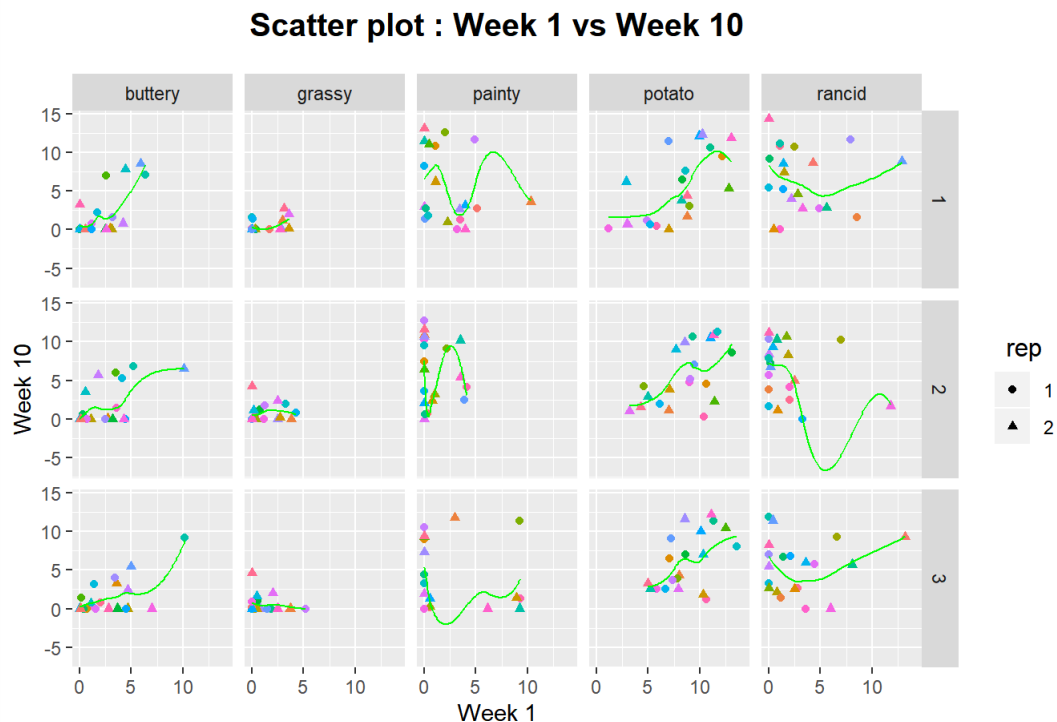

기름의 종류와 실험 반복 횟수에 따라 연구 시작 기간 1주차와 10주차 점수에 대한 산점도를 그렸다. 반복횟수와 관계없이 buttery와 potato의 경우 1주차와

10주차의 점수가 양의 상관을 가지는 것처럼 보인다. 10주차와 1주차의 양상이 초기와 비슷하다. 특히 potato의 경우 반복 횟수가 증가할수록 1주차와 10주차의 점수 분포의 위치가 높은 곳으로 이동하는 것을 볼 수 있다. grassy의 경우 반복에 상관없이 다른 기름에 비해 1주차와 10주차의 점수 분포가 낮은 곳에 밀집된다.

Your turn 2.

For this your turn use the fbiwide data.

2-1.

Use gather or pivot_longer from the tidyr package to combine the variables for the different types of crimes into one variable. Call the key-value pair "Type" and "Incidences". Compute a crime rate.

```
fbiwide <-read_rds('fbiwide.rds')

fbiwide %>%
  pivot_longer(cols = Aggravated.assault:Robbery ,
               names_to = 'type',
               values_to = 'incidences') %>%
  mutate(crime_rate = incidences/Population)-> prob2_1

knitr::kable(prob2_1 %>% head(15))
```

| State	Abb	Year	Population	type	incidences	crime_rate
Alabama	AL	1961	3302000	Aggravated.assault	NA	NA
Alabama	AL	1961	3302000	Burglary	11205	0.0033934
Alabama	AL	1961	3302000	Larceny.theft	18801	0.0056938
Alabama	AL	1961	3302000	Legacy.rape	NA	NA
Alabama	AL	1961	3302000	Motor.vehicle.theft	2535	0.0007677
Alabama	AL	1961	3302000	Murder	NA	NA
Alabama	AL	1961	3302000	Rape	NA	NA
Alabama	AL	1961	3302000	Robbery	NA	NA
Alabama	AL	1961	3302000	Aggravated.assault	4255	0.0012886
Alabama	AL	1961	3302000	Burglary	NA	NA
Alabama	AL	1961	3302000	Larceny.theft	NA	NA
Alabama	AL	1961	3302000	Legacy.rape	252	0.0000763
Alabama	AL	1961	3302000	Motor.vehicle.theft	NA	NA
Alabama	AL	1961	3302000	Murder	427	0.0001293
Alabama	AL	1961	3302000	Rape	NA	NA

2-2.

Only consider crime rates for a and Minnesota. low Use spread or pivot_wider to create incidence columns for each of these states. Plot crimes in Iowa against crimes in Minnesota, colour by type of crime. Note: you need to exclude some variables.

```
prob2_1 %>%
  filter(State %in% c('Iowa','Minnesota')) %>%
  mutate(row=rep(1:928,2)) %>%
  select(State,row,type,crime_rate) %>%
  pivot_wider(names_from = State,
              values_from = crime_rate) %>%
  ggplot(aes(x=Iowa, y=Minnesota))+
  geom_point(aes(color=type))+
  geom_smooth()+
  labs(title='IOWA vs Minnesota')+
  theme(plot.title = element_text(hjust = 0.5,size = 15, face = "bold",margin=
```

IOWA vs Minnesota
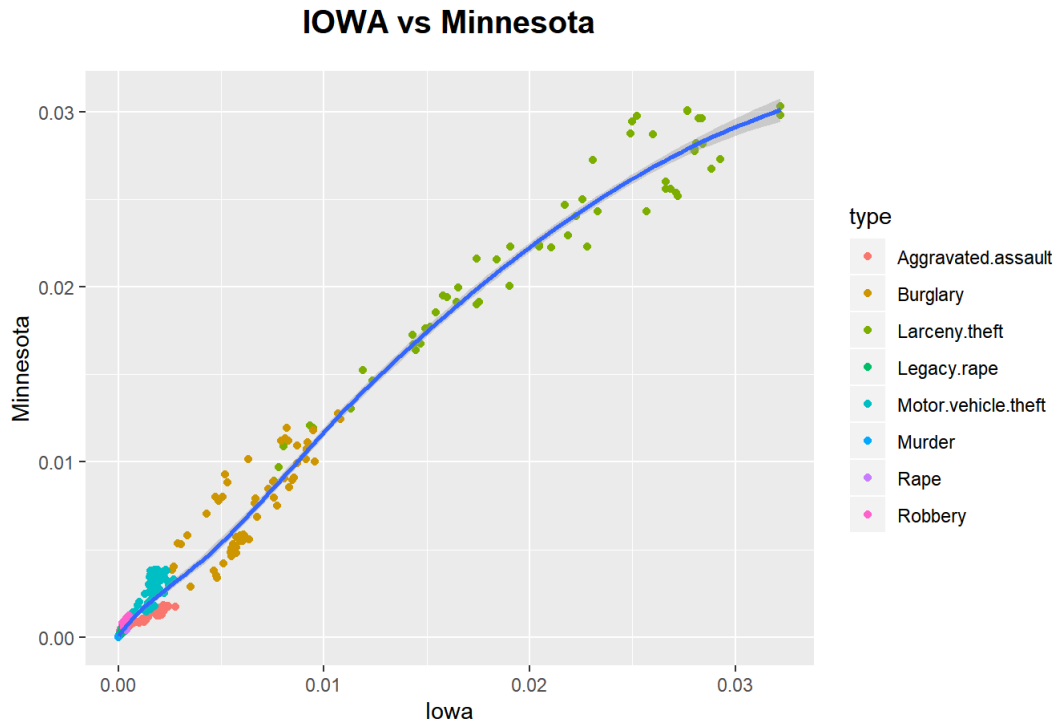
Iowa와 Minnesota의 crime rate에 대해 산점도를 그렸더니 양의 상관을 가진다. 특히, Larceny theft의 경우 두 지역에서 crime rate이 가장 높았고 Burglary, Motor.vehicle theft가 차례로 높게 나타났다. Murder의 경우 두 지역에 모두 가 장 낮게 일어나는 crime이었다.