# homework3

Hyeonho Lee

2018년 8월 23일

Exercises 8

Repeat exercise #3. This time, use the variables cohort, zygosity to facet, and use different colors to indicate the sub-groups for which the heights are significantly similar. Comment on your finding.

```
condition = twinData %>% group_by(cohort,zygosity) %>%
  do(tidy( cor.test(~ ht1 + ht2, alternative = "greater" , data = . ))) %>%
  select(cohort,zygosity,estimate)

twinData = merge(twinData, condition, by = c('cohort', 'zygosity'))

twinData %>% ggplot(mapping = aes(ht1,ht2,color=estimate)) +
  geom_point() + facet_grid(rows = vars(cohort), cols = vars(zygosity))
```
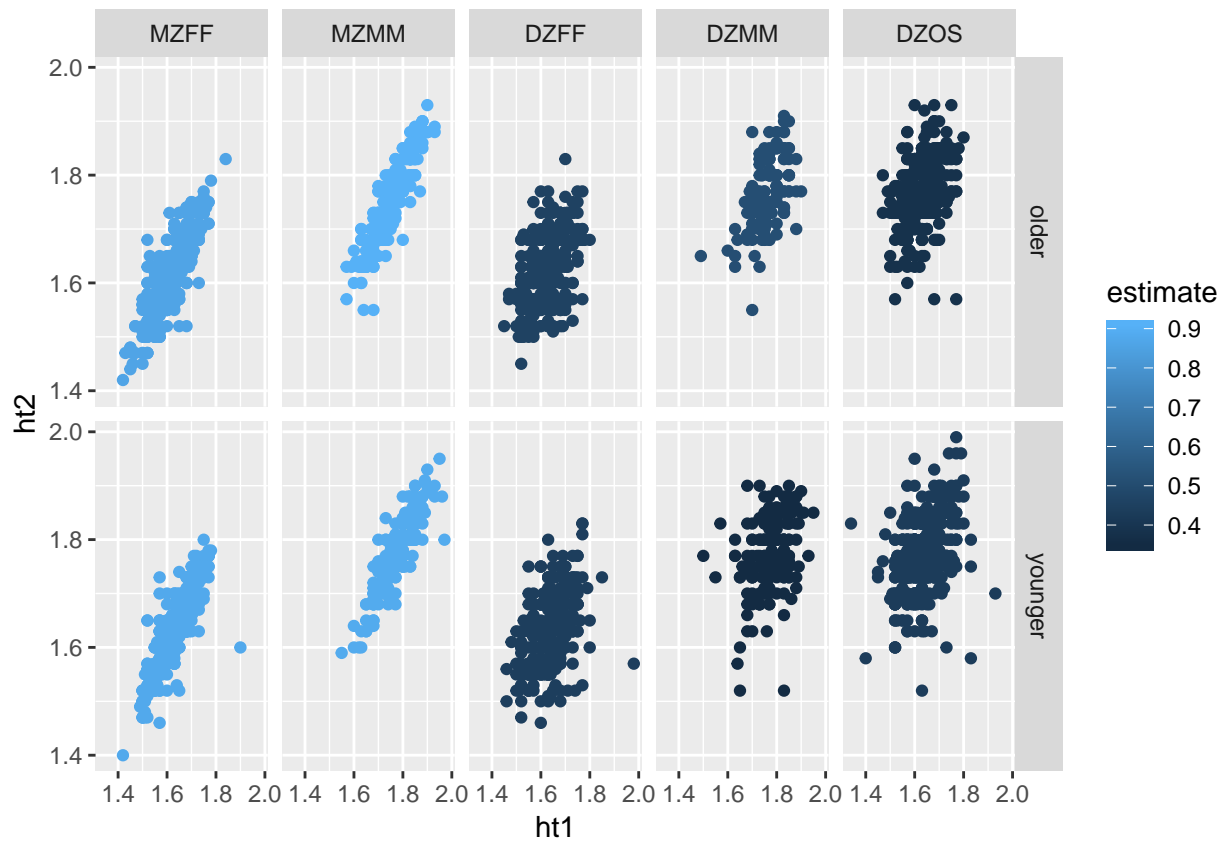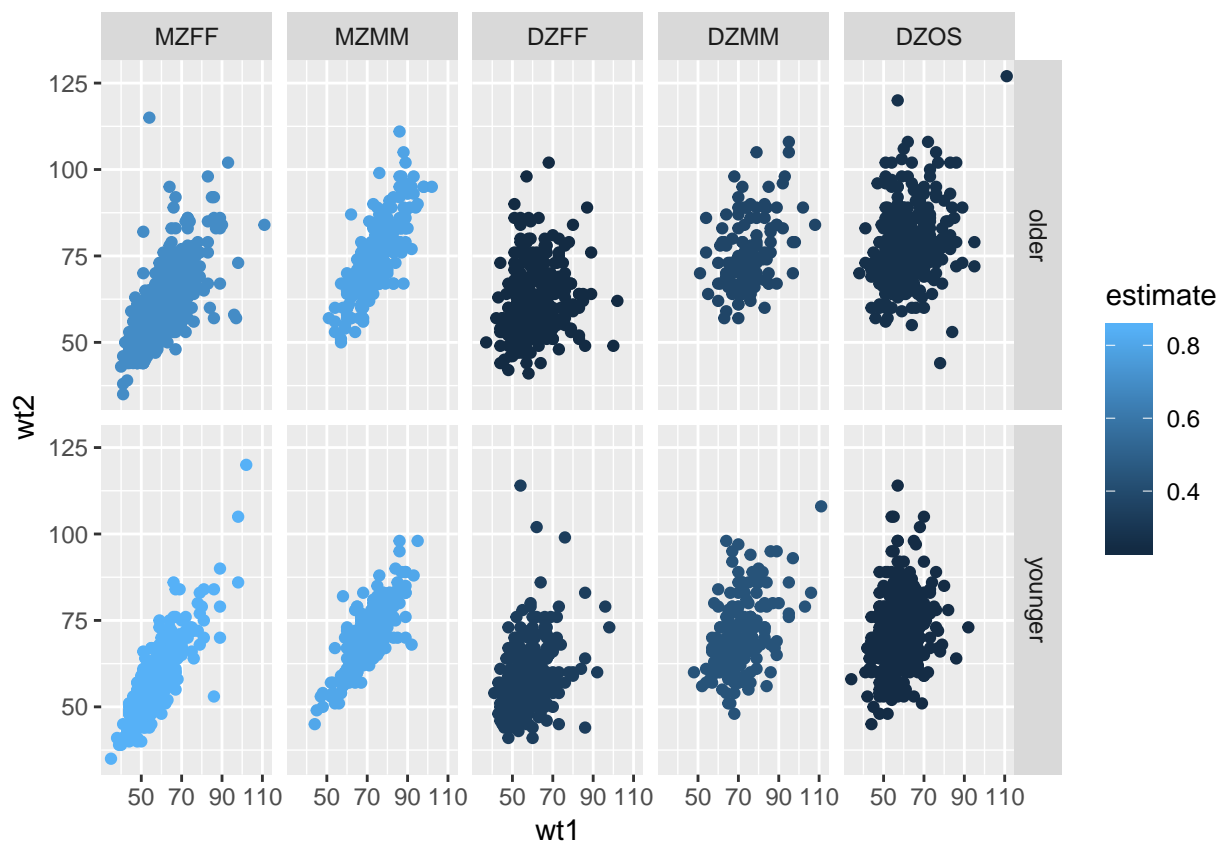
Exercises 9
Repeat exercise #8, but compare weight this time. You should be able to recycle almost all code chunks.

```
condition2 = twinData %>% group_by(cohort,zygosity) %>%
  do(tidy( cor.test(~ wt1 + wt2, alternative = "greater" , data = . ))) %>%
  select(cohort,zygosity,estimate)

twinData = twinData %>% select(-estimate) %>% merge(condition2, by = c('cohort', 'zygosity'))

twinData %>% ggplot(mapping = aes(wt1,wt2,color=estimate)) +
  geom_point() + facet_grid(rows = vars(cohort), cols = vars(zygosity))
```
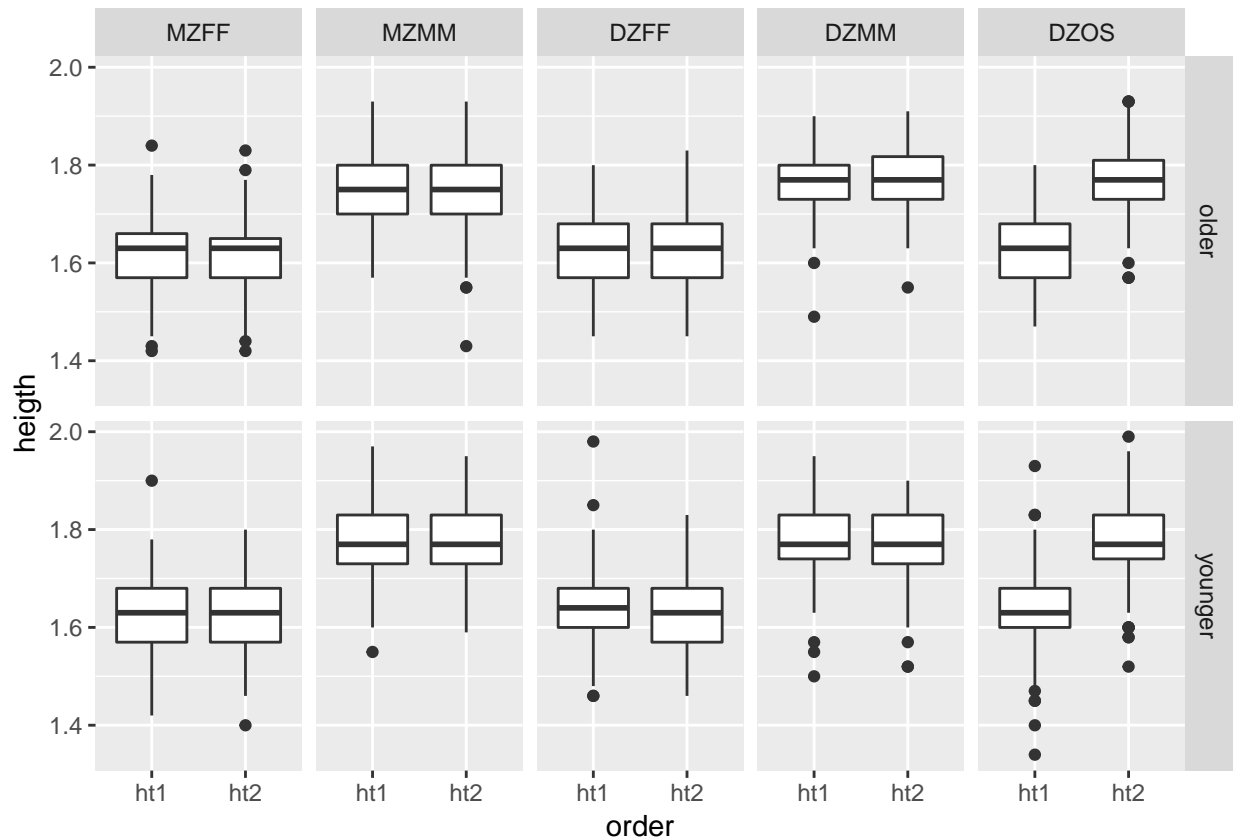
Exercises 10
Recreate the following graphic. This involves transforming twinData into a narrow form using gather(). You might want to take a look at Lecture 7 note for boxplots.

```r
twinData_1 = twinData %>% select(ht1, ht2) %>% gather(key = "order", value = "heigth") %>%
  cbind(twinData$cohort, twinData$zygosity)
colnames(twinData_1)[3:4] = c('cohort', 'zygosity')

twinData_1 %>% na.omit() %>% ggplot(mapping = aes(order, heigth)) +
  geom_boxplot() + facet_grid(rows = vars(cohort), cols = vars(zygosity))
```


Exercises 11
Inspect the data graphic. Is there any need to adjust the hypothesis (posed in Question #2)?
Are the first-born taller than the second-born?

전체 데이터를 보았을 때 상관성이 있어 보였으나, 각 범주로 나누어 비교해 본 결과, DZOS 경우에만 키의 차이를 보이는 것을 볼 수 있다.

Exercises 12

Use the paired t-test to test whether the first-born's height is significantly different from the second-born for each subgroup.

Why do we use the paired t-test, as opposed to using the two-sample t-test?

```
twinData_t.test = twinData %>% group_by(cohort, zygosity) %>%
  do(tidy(t.test(.$ht1, .$ht2, data=., paired = TRUE)))

knitr::kable(twinData_t.test[,c(1:5,9,10)], caption = 't-test')
```

Table 1: t-test

cohort	zygosity	estimate	statistic	p.value	method	alternative
older	MZFF	0.0012716	0.9528187	0.3410400	Paired t-test	two.sided
older	MZMM	-0.0002389	-0.1314739	0.8954925	Paired t-test	two.sided
older	DZFF	0.0031411	0.9159889	0.3602417	Paired t-test	two.sided
older	DZMM	-0.0063464	-1.2020046	0.2314056	Paired t-test	two.sided
older	DZOS	-0.1411766	-40.4927683	0.0000000	Paired t-test	two.sided
younger	MZFF	0.0001800	0.1280849	0.8981288	Paired t-test	two.sided
younger	MZMM	0.0012818	0.6346907	0.5261947	Paired t-test	two.sided
younger	DZFF	0.0076053	1.9417935	0.0529878	Paired t-test	two.sided
younger	DZMM	0.0021255	0.3758702	0.7074305	Paired t-test	two.sided
younger	DZOS	-0.1429043	-43.1645170	0.0000000	Paired t-test	two.sided

쌍둥이 이기 때문에 한 집단에 대한 비교로 생각 할 수 있고, 그러므로 paired 비교가 적합하다.

Test whether there is a difference with significance level 5%, for each subgroup.

```
twinData_t.test = twinData %>% group_by(cohort, zygosity) %>%
  do(tidy(t.test(.$ht1, .$ht2, data=., paired = TRUE, conf.level = 0.95)))

knitr::kable(twinData_t.test[,c(1,2,3,7,8)], caption = 'significance level 5%')
```

Table 2: significance level 5%

cohort	zygosity	estimate	conf.low	conf.high
older	MZFF	0.0012716	-0.0013490	0.0038922
older	MZMM	-0.0002389	-0.0038152	0.0033375
older	DZFF	0.0031411	-0.0036011	0.0098833
older	DZMM	-0.0063464	-0.0167857	0.0040928
older	DZOS	-0.1411766	-0.1480318	-0.1343213
younger	MZFF	0.0001800	-0.0025799	0.0029399
younger	MZMM	0.0012818	-0.0026952	0.0052587
younger	DZFF	0.0076053	-0.0000986	0.0153091
younger	DZMM	0.0021255	-0.0090286	0.0132797
younger	DZOS	-0.1429043	-0.1494094	-0.1363993

Finally, recreate the above graphic with different colors indicating the results of t-tests.

```
twinData_1 = merge(twinData_1, twinData_t.test, by = c('cohort', 'zygosity'))

twinData_1 %>% ggplot(mapping = aes(order, heigth, fill = p.value)) +
  geom_boxplot() + facet_grid(rows = vars(cohort), cols = vars(zygosity))
```