Take Home Exam project
for soil in different condition
Name: Junxi Chen email: jxichen@ucdavis.edu instructor: T.Kim

1.Introduction. State the question you are trying to answer, why it is a question of interest (why we might be interested in the answer), and what statistical technique you are going to use. This must be a non-parametric technique.

The question that I am interest Is that the different condition for gap and growth will affect the soil core's respiration, the amount of carbon dioxide. Does the group in gap condition not produce as much on average as the growth condition group? Do they have same effectiveness? the reason that I am interested in this Is because if we can find the reason to realized which condition can affect the respiration effect, the amount of carbon dioxide, we can utilize this to control the growth of the soil core and help with global environment. I am going to use mean, standard deviation, the graph, the value of test statistic, and permutation test and Wilcox test.

Part2:

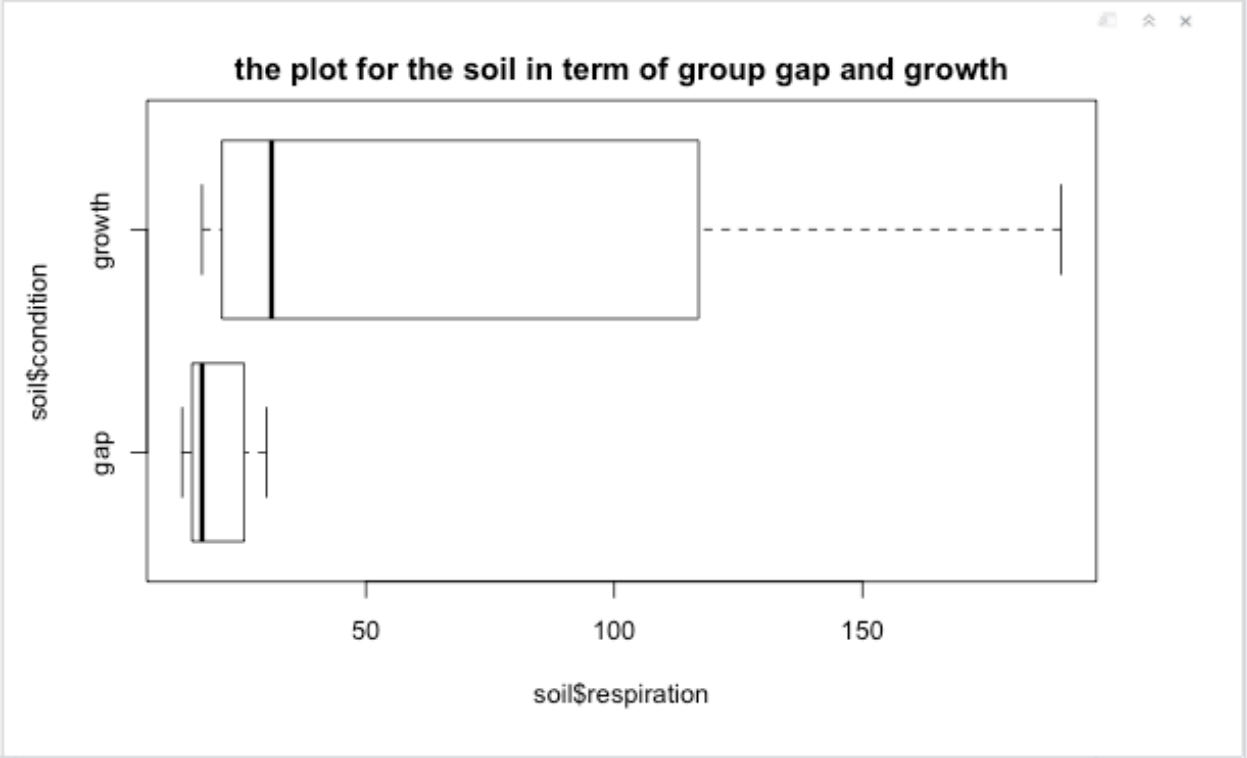
soil$condition <fctr>	soil$respiration <dbl>
gap	19.75000
growth	73.42857

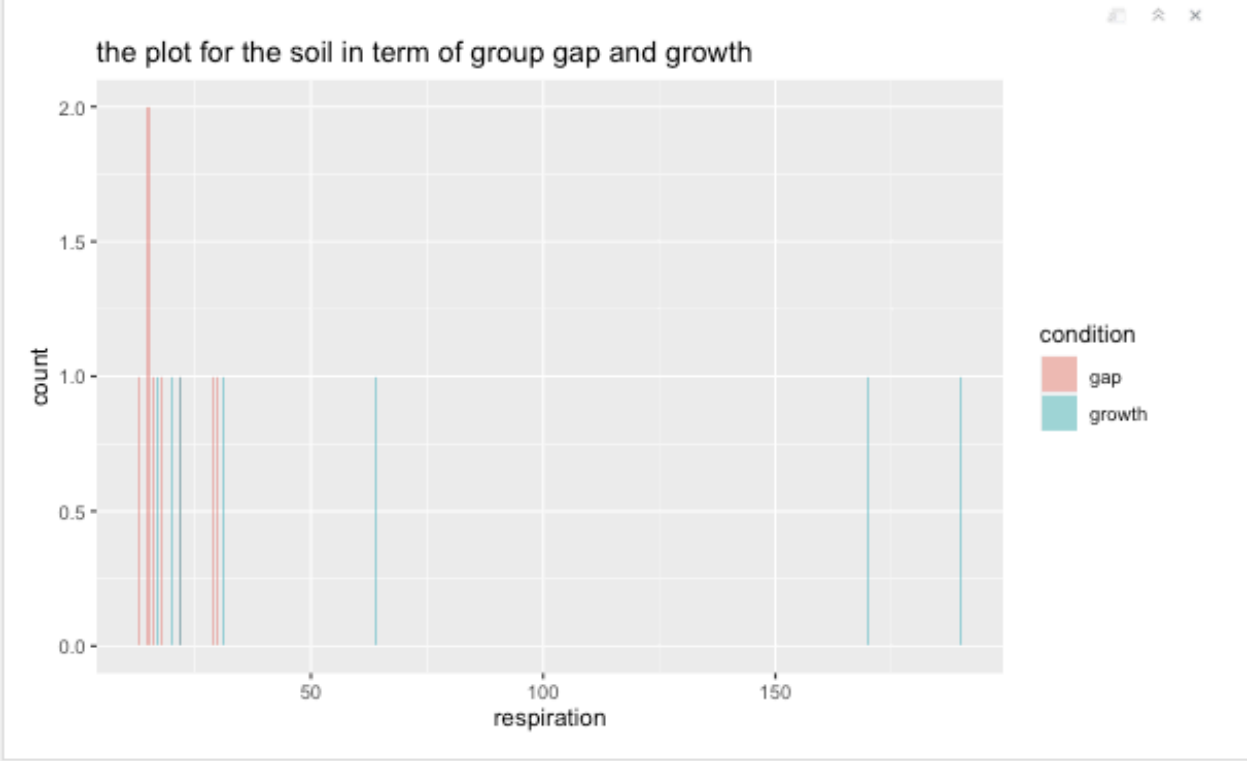
2 rows

soil$condition <fctr>	soil$respiration <dbl>
gap	6.584614
growth	74.708003

2 rows

so, the standard deviation for this is group gap is 6.584614, for growth is 74.708003. the mean for the group gap is 19.75 and for growth is 73.42857.





the interpretation for the plot is as we can see from the boxplot, there is a really big value in growth condition, which is outlier in this case, and the median is mainly located in 31, the max is 190, the min is 17. for the condition with gap, the median is mainly located in 17, the min is 13, the max is 30. and for this condition, it seems normally distributed, and not include some

outliers. And as we can see from the graph, because it appears the standard deviation for two groups are different, so we use non-parametric methods.
And the median, min and max for the two groups are

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   13.0    16.5    22.0    44.8    30.5   190.0
```

If we see in separately, for gap group

```
[1] 13.0 15.0 17.0 25.5 30.0
```

For growth group

```
[1]  17  21  31 117 190
```

Part3:
our null and alternative are: H0: F1(x) = F2(x) the respiration under the gap condition have the same amount with group under condition growth.
HA :F1(x)$\geqslant$F2(x)orμ1 <μ2, the respiration under the gap condition do not have the same amount with group under condition growth. The gap has more respiration than growth.
Because we have use R to test those data, and if we use permutation test to calculate the p value for this, we can get p value is `0.01958`, test statistic is -1.8386. and `if we want to find the confidence interval for this, we need to use equation that p +-Z*sqrt(p*(1-p)/n), so the confidence interval is 0.08535397, and upper bound is -0.04619397.`
And `if we use Wilcox to test about this, we can also get` the p value is 0.01554002, And Approximate p value is 0.01598472, the test statistic is -1.8386.

Part 4:
because we calculate the approximate p value, the absolute difference between those is really small, and so we conclude that for alpha less than 99%, we fail to reject the null hypothesis, for alpha, confidence level larger than 99%, we would to reject H0, and cannot conclude that the growth condition have larger respiration than gap condition for soil core. And we use permutation test to calculate the p value, and in the same condition, if the alpha, which is(1-confidence level) is small than `0.01958, so we fail to reject the null hypothesis, if the alpha is larger, so we can reject null hypothesis that they are in different respiration.`

Part5 :
In Summary, because we want to test about the soil respiration in different conditions, so we collect the data, and use permutation to test our hypothesis. Those two different conditions will have different effect on respiration for most of times. so, we conclude that for alpha larger than 99%, which means alpha are small than 0.01,we fail to reject the null hypothesis, for alpha, confidence level less than 99%, we would to reject H0, and cannot conclude that the growth condition has larger respiration than gap condition for soil core.

Appendix R code:

```R
soil <- read.csv("data/soil-1.csv")
summary(soil$respiration, soil$condition)
mean0= aggregate(soil$respiration~ soil$condition, data= soil, mean)
mean0
sd0= aggregate(soil$respiration~ soil$condition, data= soil, sd)
sd0
aggregate(soil$respiration~ soil$condition, data= soil, median)
boxplot(soil$respiration~soil$condition, data= soil, main="the plot for the soil in term of group gap
and growth", horizontal = TRUE)
ggplot(soil, aes(x=respiration, fill=condition)) + geom_histogram(binwidth=.5, alpha=.5,
position="identity") + ggtitle("the plot for the soil in term of group gap and growth")
summary(soil$condition)
growth1= c(17, 20, 170, 31, 22, 190, 64)
gap1= c(22, 29,13, 16,15, 18,15,30)
fivenum(gap1)
fivenum(growth1)
```

```{R}
library(coin)
test1= oneway_test(soil$respiration~soil$condition, data= soil, alternative= "two.sided",
distribution="exact" )
test1#so the p value is 0.01958, the test statistic is -1.8386

a=1-0.95
qnorm(1- a/2)
n=15

exac.pval= pvalue(wilcox_test(soil$respiration~soil$condition, soil, distribution ="exact", alternative=
"less"))
approx.pval=  pvalue(wilcox_test(soil$respiration~soil$condition, soil, alternative= "less"))
exac.pval#0.01554002
approx.pval#0.01598472
exac.pval- approx.pval#<2.220446e-16, so this is really small.

LB11=0.01958 + 1.8386*sqrt(0.01958 *(1-0.01958 )/ n)
UB1= 0.01958 - 1.8386*sqrt(0.01958 *(1-0.01958 )/ n)
```