The following questions are to help you verify that you studied all of the material in the **t Tests** portion of the **Statistics Notebook**.

Match each t Tests with its corresponding hypotheses.

[1] Paired Samples t Test [3] Independent Samples t Test [2] One Sample t Test

1. $$H_0 : \mu_{\text{post - pre}} = 0$$ $$H_a : \mu_{\text{post - pre}} \neq 0$$

2. $$H_0 : \mu = 5.8$$ $$H_a : \mu \neq 5.8$$

3. $$H_0 : \mu_{\text{Group 1}} - \mu_{\text{Group 2}} = 0$$ $$H_a : \mu_{\text{Group 1}} - \mu_{\text{Group 2}} \neq 0$$

Which t Test requires that the sampling distribution of the sample mean of the differences is normally distributed?

- ( ) One Sample t Test
- (•) Paired Samples t Test
- ( ) Independent Samples t Test

Use a subset of the CO2 dataset in R to answer the following questions.

Filter the dataset so that you are considering only the *chilled* plants where conc = 250.

(Hint: Treatment == "chilled" & conc == 250)

```
> ?CO2
> View(CO2)
> library(tidyverse) #loads the filter(...) function
> CO2.chilled.250 <- filter( ... complete this command on your own in R ... )
```

What are the mean uptake values for each Type of plant (Quebec and Mississippi) for plants that are chilled at a concentration of 250?

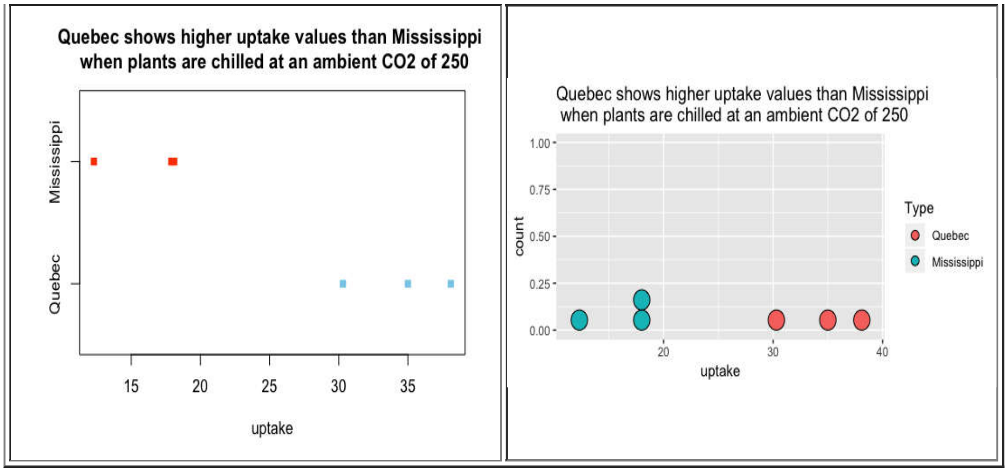
Quebec mean uptake: [34.467]

Mississipp mean uptakei: [16.100]

A side-by-side [dot plot] is best for presenting this data visually because of the small sample sizes.

Recreate **one** of these dot plots in R using either **Base** or **ggplot2** graphics.

Base R	ggplot2



Once you recerate one of the graphs above, select the graph you successfully recreated:

ggplo2 dot plot using geom_dotplot(...) DON'T select this unless you have actually successfully recreated one of the graphs.

These two sample means are clearly different. However, these results are based on a small sample of Mississippi and Quebec plants. The question that remains is whether or not this pattern holds for ALL Mississippi and Quebec plants, or if it was just a figment of this sample of data.

To determine if we should conclude that this pattern holds for the full population, conduct an appropriate hypothesis test for the hypotheses:

$$H_0 : \mu_{Miss} - \mu_{Queb} = 0$$
$$H_a : \mu_{Miss} - \mu_{Queb} \neq 0$$

where μ_{Miss} represents the (unknown) true mean CO2 uptake of ALL Mississippi plants that are chilled at a concentration of 250 and μ_{Queb} represents the true (unkown) mean CO2 uptake of ALL Quebec plants that are chilled at a concentration of 250.

Report the test statistic, the parametric distribution being used for the test statistic (including degrees of freedom if appropriate), and p-value of the test.

Test Statistic: t = 6.2075

Parametric Distribution for the Test Statistic: t distribution with 3.8818 degrees of freedom.

P-value = .003782

This shows there is sufficient evidence to conclude that Quebec plants (that are chilled at a concentration of 250) truly do have a different mean uptake than Mississippi plants (that are chilled at a concentration of 250). In other words, it is safe to conclude that the pattern we have seen in the sample data holds for the full population. Note that since our sample data shows that Quebec has a higher mean than Mississippi, and the p-value of the two-sided test is signficant, we can actually conclude that Quebic plants have higher CO2 uptake than Mississippi plants, on average.

However, before reporting our results, it is important to verify that the requirements of this hypothesis test were satisfied. Which of the following requirements should be verified for this particular hypothesis test?

- [x] The sample of Mississippi and Quebec plants need to be representative of the population. Hopefully they were selected randomly so that this would be satisfied.
- [ ] The sampling distribution of the sample mean of the differences $\bar{d}$ can be assumed to be normal. (This requirement can be assumed to be satisfied when (a) the differences themselves can be assumed to be normal, or (b) when the sample size n of the differences is large.)
- [ ] The sampling distribution of the sample mean $\bar{x}$ can be assumed to be normal. This is a safe assumption when either (a) the population data can be assumed to be normally distributed or (b) the size of the sample taken from the population is large.
- [x] The sampling distribution of the difference of the sample means $\bar{x}_1 - \bar{x_2}$ can be assumed to be normal. (This is a safe assumption when the sample size of each group is 30 or greater or when the population data from each group can be assumed to be normal.)

How many plants of each Type (Mississippi and Quebec) are in the sample of data?

Mississippi sample size: 3

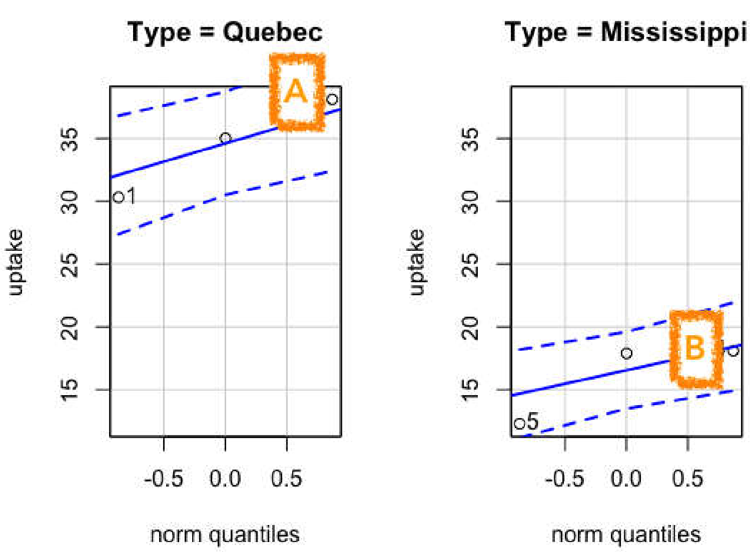
Quebec sample size: 3

These are small sample sizes. You will need to create Q-Q Plots of each sample of data to determine if the data in each group is normal or not. It turns out that making these Q-Q Plots in R is a little cumbersome. Here is the code that does it correctly.

```
> library(car)
> qqPlot(uptake ~ Type, data = CO2.chilled.250)
```

Have you successfully run the codes above in your R Console? If so, your plots should look something like the following.

To prove you have made these two Q-Q Plots, state the values of A = 3 and B = 4 .



- (•) I successfully reproduced these Q-Q Plots in my R Console.
- ( ) I have not been able to recreate these plots.

Note that the sample sizes are very small (n=3 in each group). However, normality does not appear to be violated as none of the points go outside of the red dashed lines. Thus, the requirements of the hypothesis test appear to be satisfied for these data. We are safe to report the results of our hypothesis test.

Read the "Overview" page of your "Independent Samples t Test". What are the two requirements that need to be verified before we can trust the p-value from an Independent Samples t Test?

- [x] The sampling distribution of the difference of the sample means ($\bar{x}_1 - \bar{x}_2$) *can be assumed to be normal*. (This is a safe assumption when the sample size of each group is 30 or greater or when the population data from each group can be assumed to be normal with a Q-Q Plot.)
- [ ] The sampling distribution of the sample mean of the differences $\bar{d}$ (the sample mean of the differences) can be assumed to be normal. (This second requirement can be assumed to be satisfied when (a) the differences themselves can be assumed to be normal from a Q-Q Plot, or (b) when the sample size n of the differences is large.)
- [ ] The sampling distribution of the sample mean $\bar{x}$ *can be assumed to be normal*. This is a safe assumption when either (a) the population data can be assumed to be normally distributed using a Q-Q Plot or (b) the size of the sample (n) that was taken from the population is large (at least n > 30, but "large" really depends on how badly the data is skewed).
- [x] Both samples are **representative** of the population. (Simple random samples are the best way to do this.)
- [ ] The sample of differences is representative of the population differences.
- [ ] The sample is **representative** of the population. (Having a simple random sample is the best way to do this.)

Thus, the two main things you need to check to verify that an Independent Samples t Test is appropriate are:
(Mark all that apply.)

- [x] The sample size of each group.
- [x] The normality of the data in each group using a Q-Q Plot for each group's data.