Day 4: Null hypothesis and statistical tests

How are you feeling today?

Have you heard of a null hypothesis?

Yes, and I could explain what it is

0

Yes, but I couldn't explain what it is

0

No, I haven't heard of a null hypothesis

0

Before you begin an experiment or analysis you should have two hypotheses:

- **Null hypothesis:** There is not a significant relationship between the things you are comparing.

- **Alternative hypothesis:** There is a significant relationship between the things you are comparing.

Example: You wonder whether average temperatures vary between months.

- Null hypothesis: There is no relationship between month and temperature.

- Alternative hypothesis: Summer months (June, July, August) have higher average temperatures than winter months (December, January, February).

Types of statistical tests you may use

- **T-test:** Compare means of two small sample sets.
- **Wilcoxon's signed rank test:** Alternative to t-test when data is not normally distributed.
- **Shapiro-Wilks test:** Tests data for normal distribution.
- **F-test:** Compare variances of sample sets.
- **Fisher's Exact test:** Analyze contingency tables with small sample size.
- **Chi-squared test:** Analyze contingency tables with large sample size.
- **Pearson's correlation:** Normalized measure of covariance between variables.
- **ANOVA:** Compare means of more than two groups.

These tests will give you a p-value

- P-values < 0.05 are usually considered significant, meaning you can reject the null hypothesis, though some may argue p should be less than 0.01 or even 0.001.

What is a p-value?

Nobody has responded yet.

Hang tight! Responses are coming in.
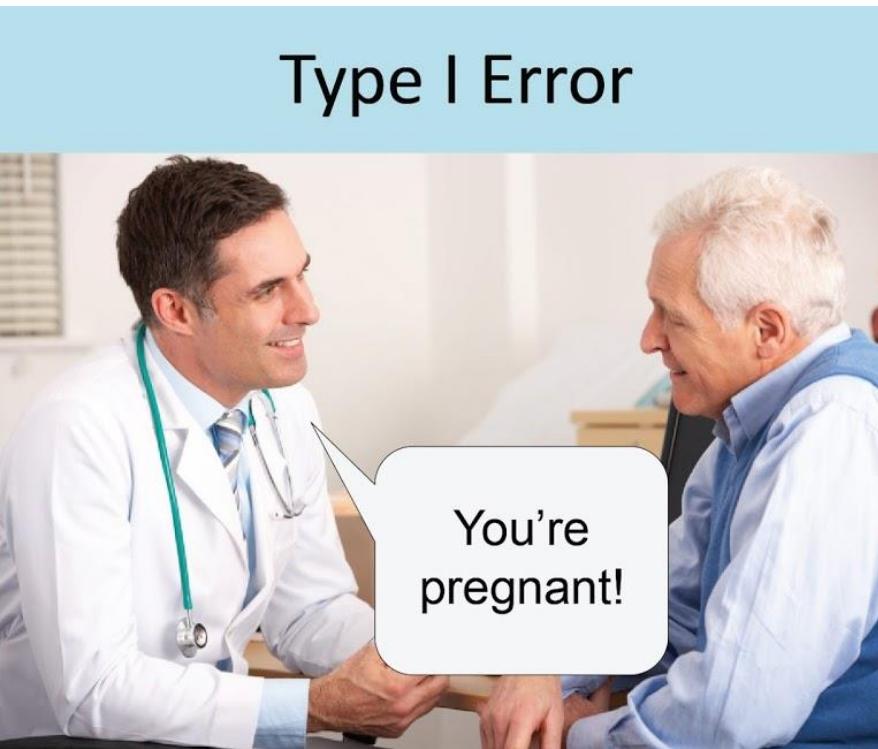
"If there were actually no effect (if the true difference between means were zero) then the probability of observing a value for the difference equal to, or greater than, that actually observed is called the p-value. In other words the p-value is the chance of seeing a difference at least as big as we have done, if, in fact, there were no real effect."

Breakdown:

If we get a significant effect (reject the null hypothesis, see a difference)...

but the null hypothesis was actually true...

what is the chance that we would see an effect equal to or bigger than what we saw when we ran this test if we run it again?

P-values can be complicated….

- For one, you may have Type I or Type II errors
- Type I: False positive, reject the null when it's true
- Type II: False negative, fail to reject the null when it is false

```{r}
library(tidyverse)
library(ggpubr)
library(rstatix)


data1 <- CO2 #First we load a dataset to work with

```

Use the data tidying skills we discussed yesterday in the example code below

```{r}
data1 %>%
  drop_na() %>%
  group_by(Treatment) %>%
  get_summary_stats(uptake, type = "mean_sd")
```

A tibble: 2 × 5

Treatment	variable	n	mean	sd
<fctr>	<fctr>	<dbl>	<dbl>	<dbl>
nonchilled	uptake	42	30.643	9.705
chilled	uptake	42	23.783	10.884

F-test: Compare variances of sample sets.

```{r}
var.test(uptake ~ Treatment, data = data1) #An F-test can be run with the function var.test(). When we run this test
looking at differences in CO2 uptake variance between treated and control groups of plants, we get a p-value > 0.05,
meaning we cannot reject the null hypothesis that the variances are equal between chilled and nonchilled plants.
```

```
        F test to compare two variances

data:  uptake by Treatment
F = 0.79504, num df = 41, denom df = 41, p-value = 0.466
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4273524 1.4790775
sample estimates:
ratio of variances
        0.7950392
```

Shapiro-Wilks test: Tests data for normal distribution

```{r}
shapiro.test(data1$uptake) #The Shapiro-Wilk test is to test for normality. Many statistical tests assume that data is
normally distributed. Our results indicate it is not normally distributed.
```

        Shapiro-Wilk normality test

data: data1$uptake
W = 0.94105, p-value = 0.0007908

T-test: Compare means of two small sample sets

Even though our data is not normal, let's go through the steps of running a t-test. First, a t-test compares the means of two groups. In this case the groups we want to compare are chilled vs. nonchilled plants. Use the code below to separate the chilled and nonchilled data and run a two-tailed t-test.

```{r}
chilled_plants <- data1 %>%
  filter(Treatment == "chilled") %>%
  pull(uptake)
unchilled_plants <- data1 %>%
  filter(Treatment == "nonchilled") %>%
  pull(uptake)
CO2_uptake_dif_t_test <- t.test(chilled_plants, unchilled_plants)
CO2_uptake_dif_t_test
```

```
        Welch Two Sample t-test

data:  chilled_plants and unchilled_plants
t = -3.0485, df = 80.945, p-value = 0.003107
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.336682  -2.382366
sample estimates:
mean of x mean of y
 23.78333   30.64286
```

Wilcoxon's signed rank test:

Since the data is not normal, we can use a Wilcoxon test to see if there is a significant difference between the groups.

```{r}
wilcox.test(uptake ~ Treatment, data = data1)
```

Warning: cannot compute exact p-value with ties
		Wilcoxon rank sum test with continuity correction

data: uptake by Treatment
W = 1187.5, p-value = 0.006358
alternative hypothesis: true location shift is not equal to 0

To plot statistical test results on a bar plot, they need to be in a certain format. We can get this format easily using the compare_means() function.

```{r}
t.test <- compare_means(
  uptake ~ Treatment,
  data = data1,
  method = "t.test"
)

t.test

```
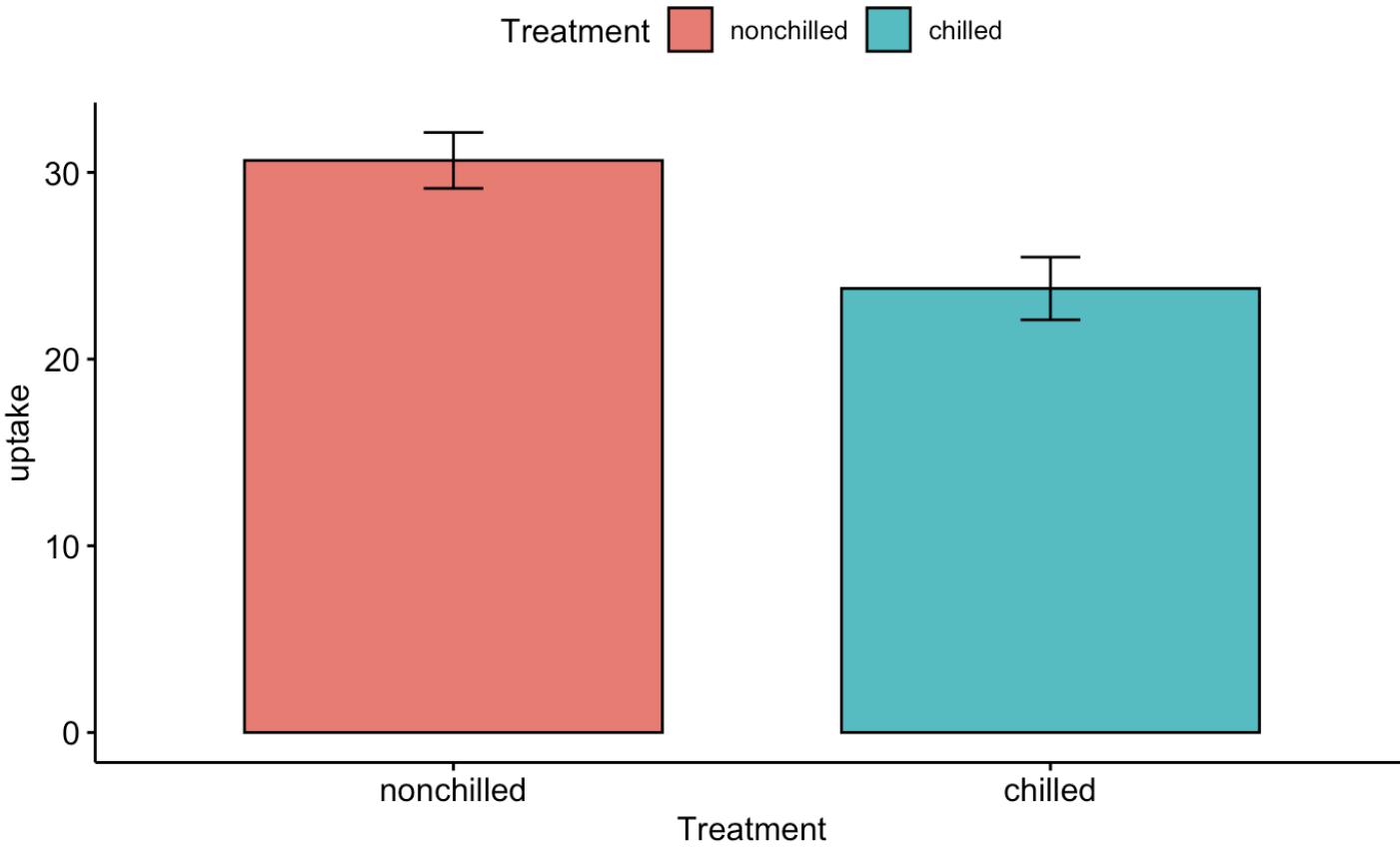
A tibble: 1 × 8
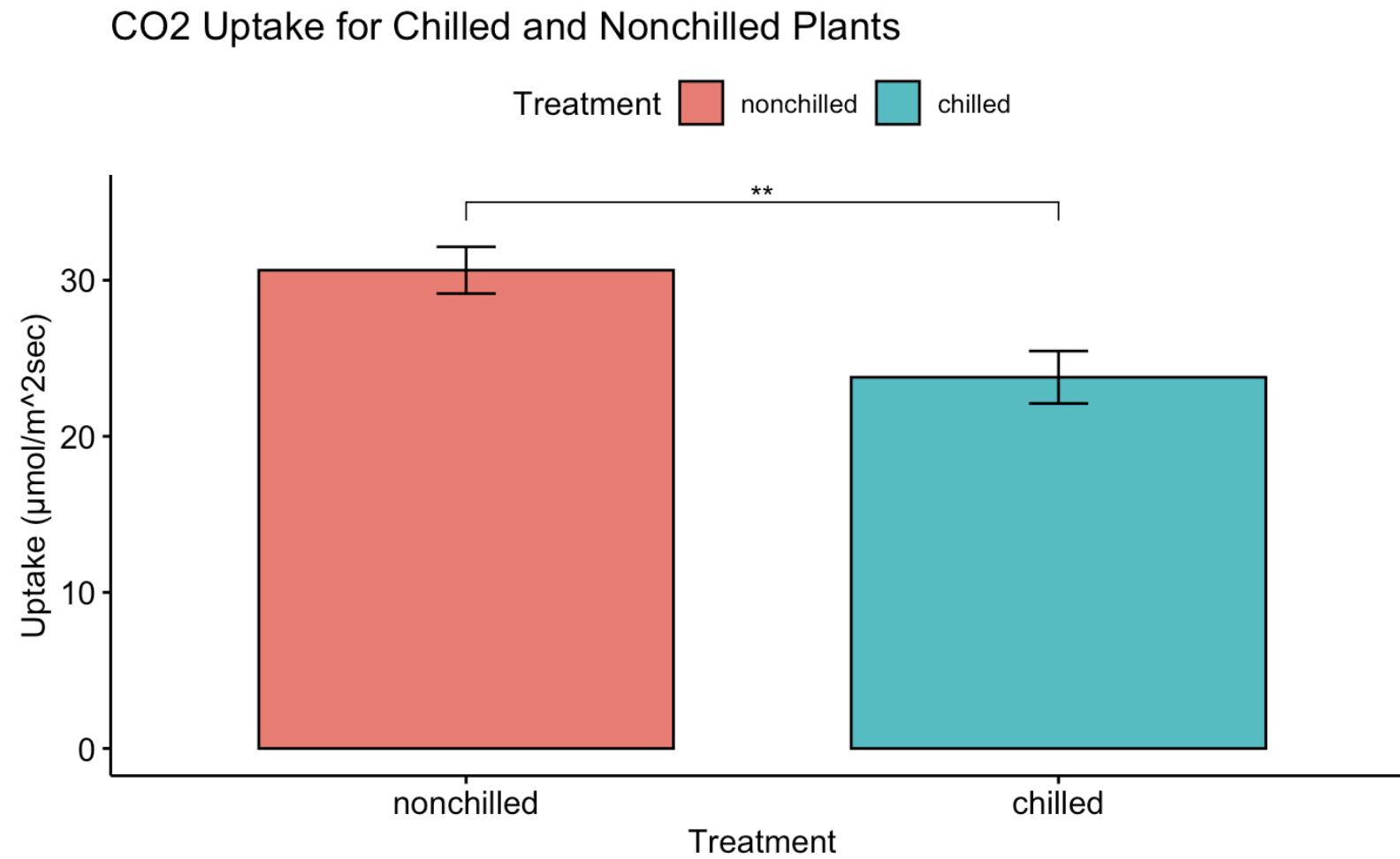
.y. <chr>	group1 <chr>	group2 <chr>	p <dbl>	p.adj <dbl>	p.format <chr>	p.signif <chr>	method <chr>
uptake	nonchilled	chilled	0.003106937	0.0031	0.0031	**	T-test

1 row

```{r}
p <- ggbarplot(
  data = data1,
  x = "Treatment",
  y = "uptake",
  fill = "Treatment",
  add = "mean_se",
  color = "black"
)
p
```

```{r}
p + stat_pvalue_manual(
  t.test,
  y.position = 35,
  label = "p.signif"
) + ylab("Uptake (µmol/m^2sec)") + ggtitle("CO2 Uptake for Chilled and Nonchilled Plants")
```


CO2 Uptake for Chilled and Nonchilled Plants

Find whether there is a significant difference between plants from Quebec and plants from Mississippi. Make sure to check normality. If the data is not normal use a Wilcoxon test. If it is normal use a t-test.

- What is your first step to investigate differences between plants from Quebec and plants from Mississippi?

Find whether there is a significant difference between plants from Quebec and plants from Mississippi. Make sure to check normality. If the data is not normal use a Wilcoxon test. If it is normal use a t-test.

```{r}
data1 %>%
  drop_na() %>%
  group_by(Type) %>%
  get_summary_stats(uptake, type = "mean_sd")
```
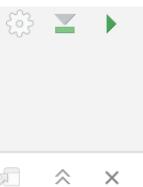
A tibble: 2 × 5

Type	variable	n	mean	sd
<fctr>	<fctr>	<dbl>	<dbl>	<dbl>
Quebec	uptake	42	33.543	9.674
Mississippi	uptake	42	20.883	7.816

2 rows

```{r}
var.test(uptake ~ Type, data = data1)
shapiro.test(data1$uptake)
```
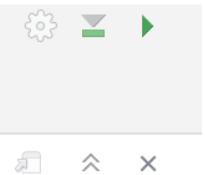
```
        F test to compare two variances

data:  uptake by Type
F = 1.532, num df = 41, denom df = 41, p-value = 0.1763
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.8234755 2.8500701
sample estimates:
ratio of variances
          1.53198


        Shapiro-Wilk normality test

data:  data1$uptake
W = 0.94105, p-value = 0.0007908
```

```{r}
wilcox.test(uptake ~ Type, data = data1)
```
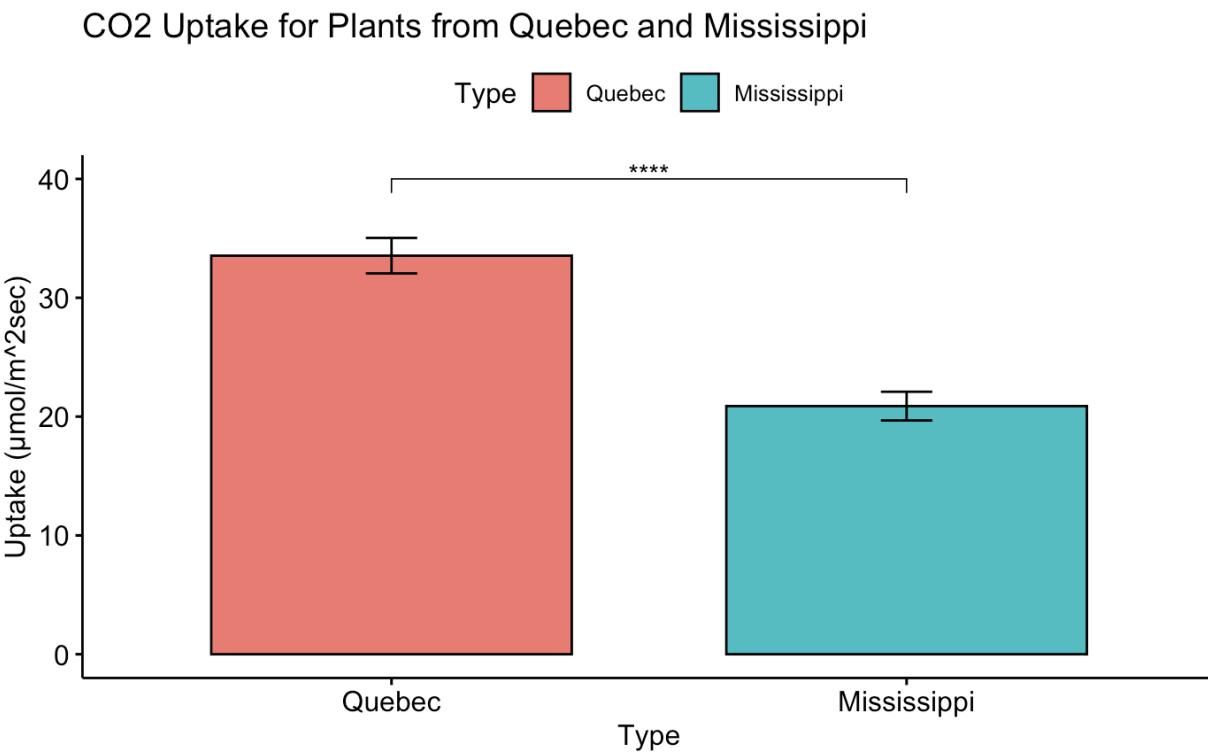
```
Warning: cannot compute exact p-value with ties
        Wilcoxon rank sum test with continuity correction

data:  uptake by Type
W = 1489, p-value = 5.759e-08
alternative hypothesis: true location shift is not equal to 0
```

```{r}
wilcox.test <- compare_means(
  uptake ~ Type,
  data = data1,
  method = "wilcox.test"
)

wilcox.test

p + stat_pvalue_manual(
  wilcox.test,
  y.position = 40,
  label = "p.signif"
) + ylab("Uptake (µmol/m^2sec)") + ggtitle("CO2 Uptake for Plants from Quebec and Mississippi")
```


CO2 Uptake for Plants from Quebec and Mississippi

What do you hypothesize for plant treatment and CO_2 uptake from plants from Quebec and plants from Mississippi?

0

Nobody has responded yet.

Hang tight! Responses are coming in.

Which of the following can be inferred, checking your p-values after grouping your dataset by the "Type" category?

Chilling plants affects CO2 uptake in plants from Quebec

0

Chilling plants affects CO2 uptake in plants from Mississippi

0

Chilling plants affects CO2 uptake in plants from Quebec and Mississippi

0

Chilling plants does not affect CO2 uptake in Quebec or Mississippi

0

```r
separated.stats <- data1 %>%
  group_by(Type) %>% # Group  the data by where the plants came from first
    wilcox_test(uptake ~ Treatment) %>% # Within each type of plant (plant from Quebec or plant from Mississippi) perform
a Wilcoxon test to look for diffferences in CO2 uptakes between treatments
  adjust_pvalue() %>% # Adjust p-value for multiple comparisons. For more about why we adjust p-values, check out the
dead salmon paper on Github for a fun read!
  add_significance()
separated.stats
```
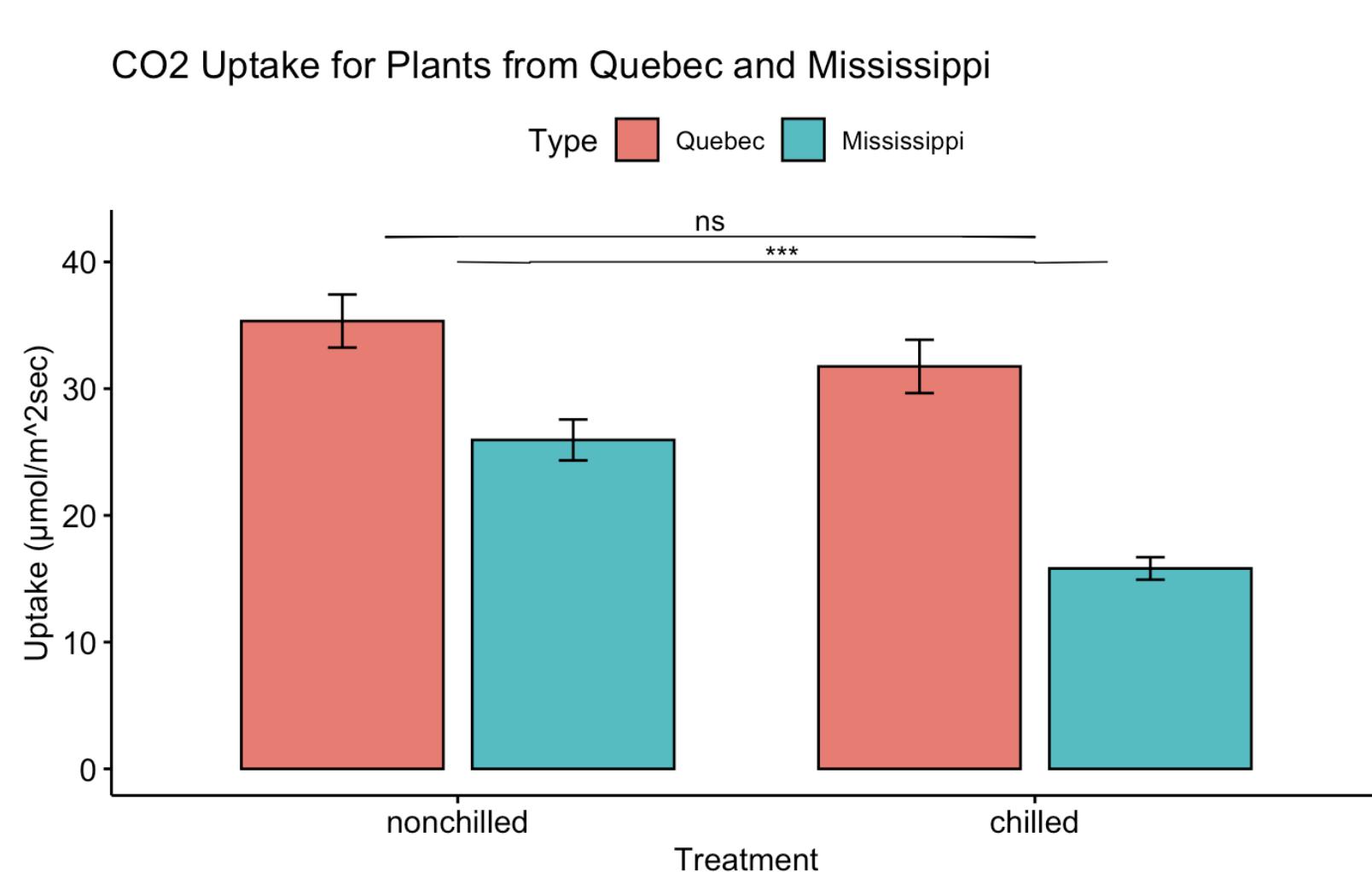
A tibble: 2 × 10

Type <fctr>	.y. <chr>	group1 <chr>	group2 <chr>	n1 <int>	n2 <int>	statistic <dbl>	p <dbl>	p.adj <dbl>	p.adj.signif <chr>
Quebec	uptake	nonchilled	chilled	21	21	290.0	0.082600	0.082600	ns
Mississippi	uptake	nonchilled	chilled	21	21	378.5	0.000074	0.000148	***

2 rows

```{r}
q + stat_pvalue_manual(
  separated.stats,
  y.position = c(42,40),
  label = "p.adj.signif",
  position = position_dodge(0.5),
  tip.length = 0.002
) + ylab("Uptake (μmol/m^2sec)") + ggtitle("CO2 Uptake for Plants from Quebec and Mississippi")
```


CO2 Uptake for Plants from Quebec and Mississippi

Which of the following applies to you?

I have a sibling, I'm right handed

0%

I have a sibling, I'm left handed

0%

I don't have a sibling, I'm right handed

0%

I don't have a sibling, I'm left handed

0%