Continuous y and Categorical x Data

Group 2: Gabriella Dauber, Isabelle DuMontier, and Alan Lu

Data Types

Continuous

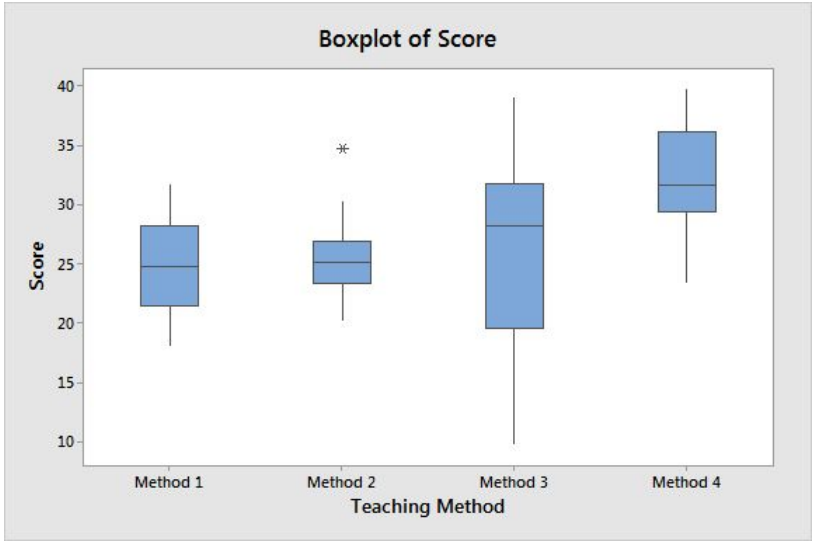
- Variables that exist on a continuous scale
- Can exist at any point in the variable's range
- Often represented as numerical data class in R

Categorical

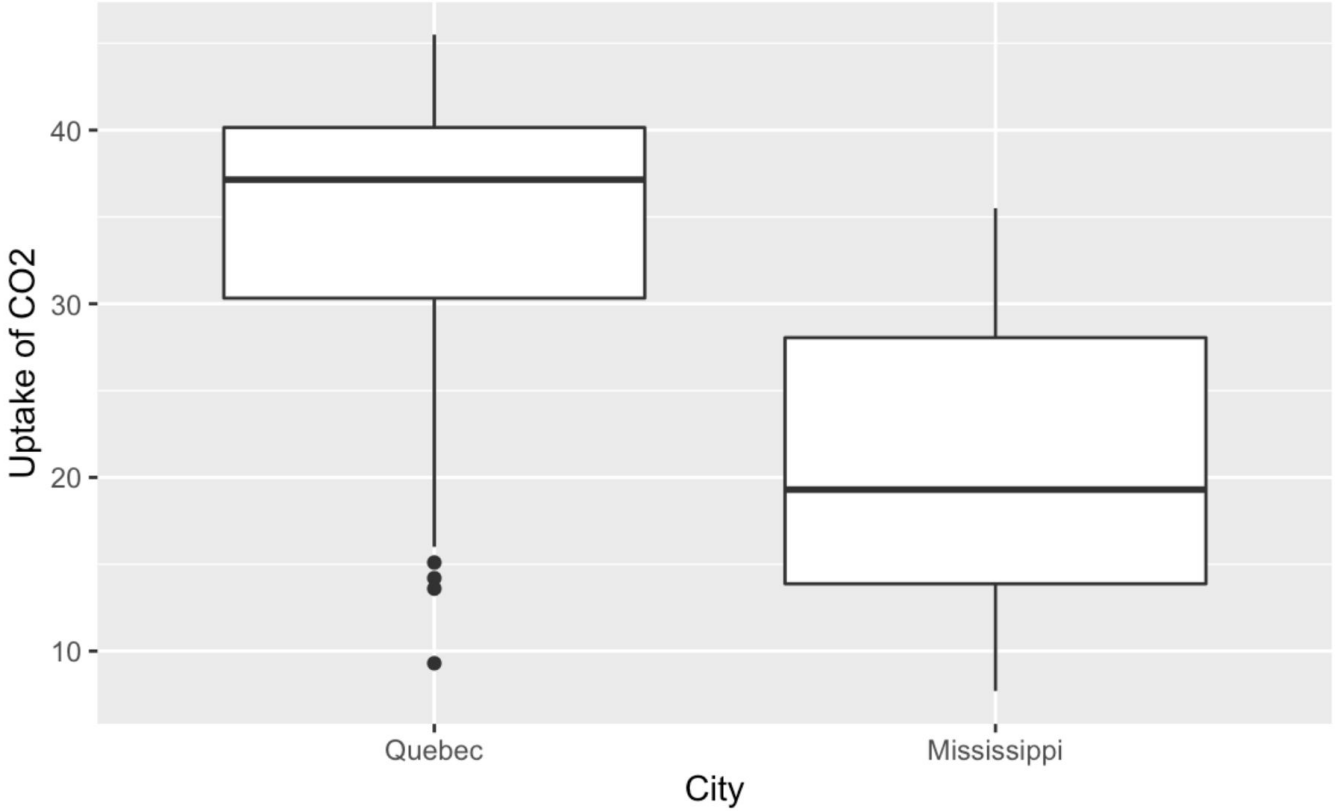
- Variables that are non-numerical and form discrete groups
- Often represented as data class factor in R

Box Plots

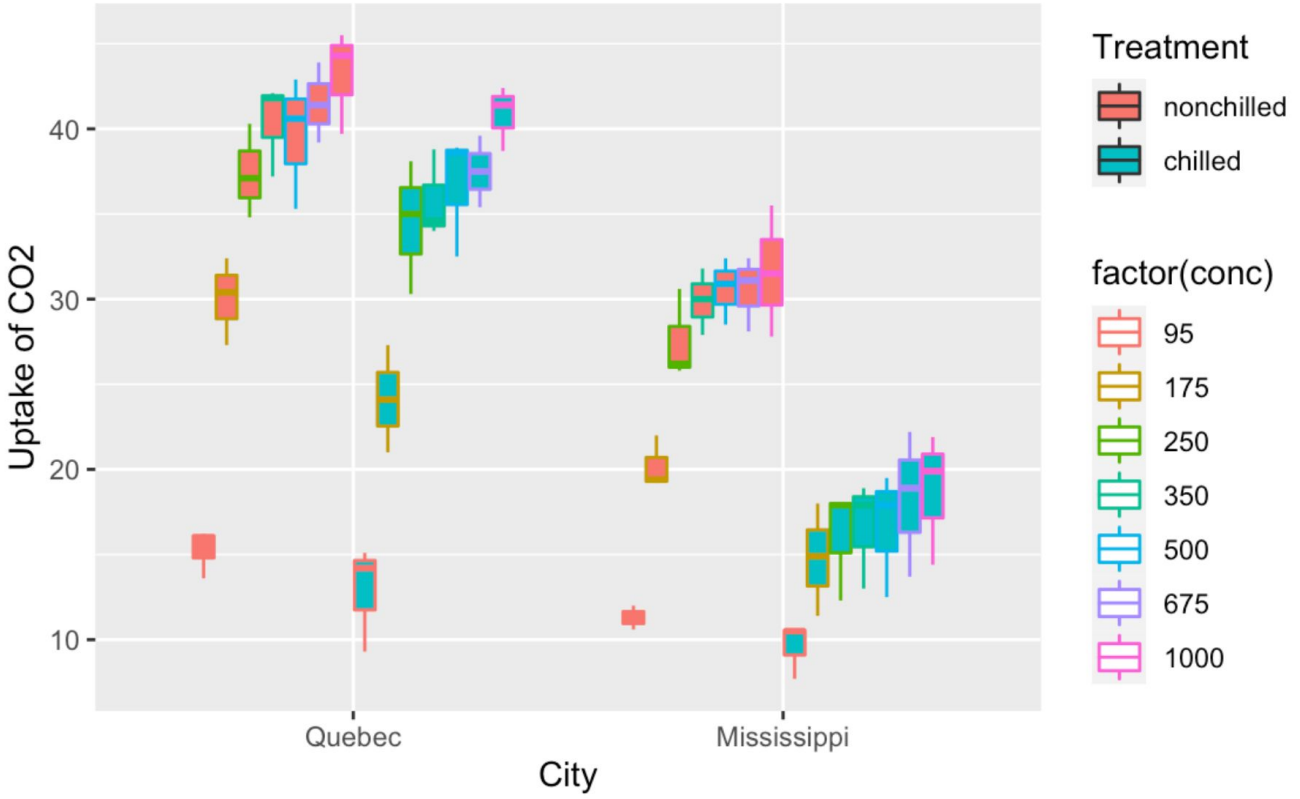
- Method for graphically depicting groups of numerical data through their quartiles
- They also may have lines extending from the boxes indicating variability outside the upper and lower quartiles
- The next few slides show examples of the possible ways of producing a box plot for the data set CO2 in R



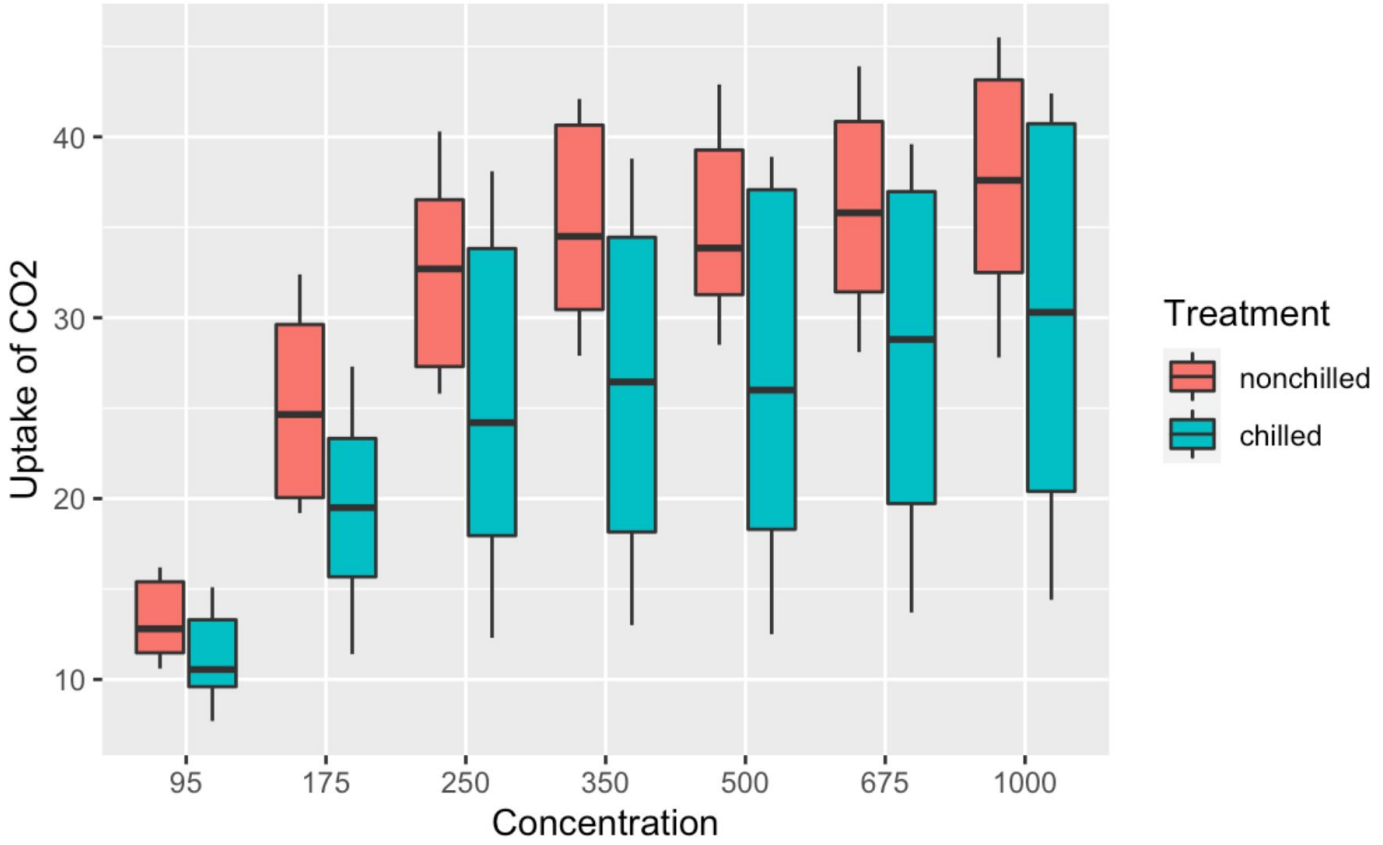
```
data(CO2)
library(ggplot2)
ggplot(CO2, aes(x=Type, y=uptake)) + geom_boxplot() + xlab("City")+ ylab("Uptake of CO2")
```



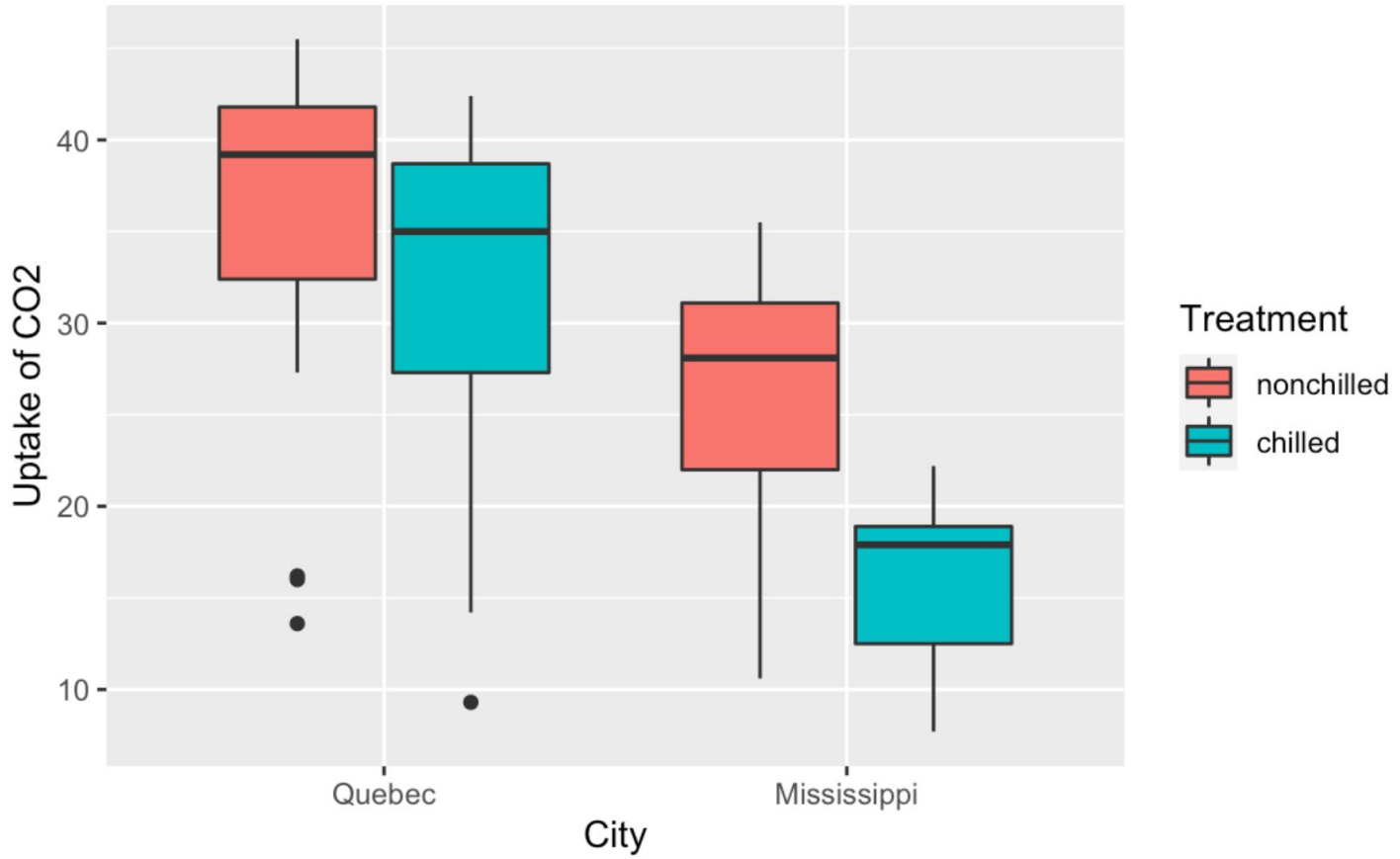
```
ggplot(CO2, aes(x=Type, fill=Treatment, y=uptake, color=factor(conc))) + geom_boxplot() + xlab("City") + ylab("Up
take of CO2")
```



```
ggplot(CO2, aes(x=factor(conc), fill=Treatment, y=uptake)) + geom_boxplot() + xlab("Concentration") + ylab("Uptak
e of CO2")
```

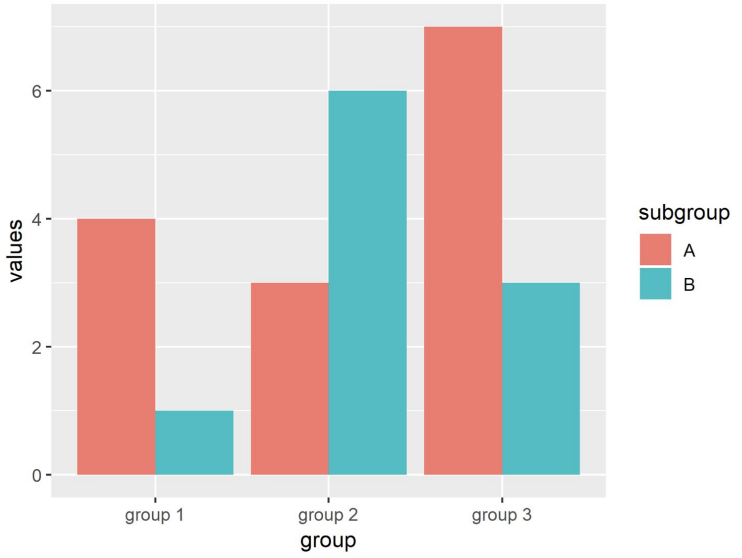


```
ggplot(CO2, aes(x=Type, fill=Treatment, y=uptake)) + geom_boxplot() + xlab("City") + ylab("Uptake of CO2")
```



Bar Plots

- Method to graphically compare the numbers/frequencies of a category or characteristic with another
- Bars break down categorical data by group
- The next few slides show examples of the possible ways of producing bar graphs for the data set CO2 in R

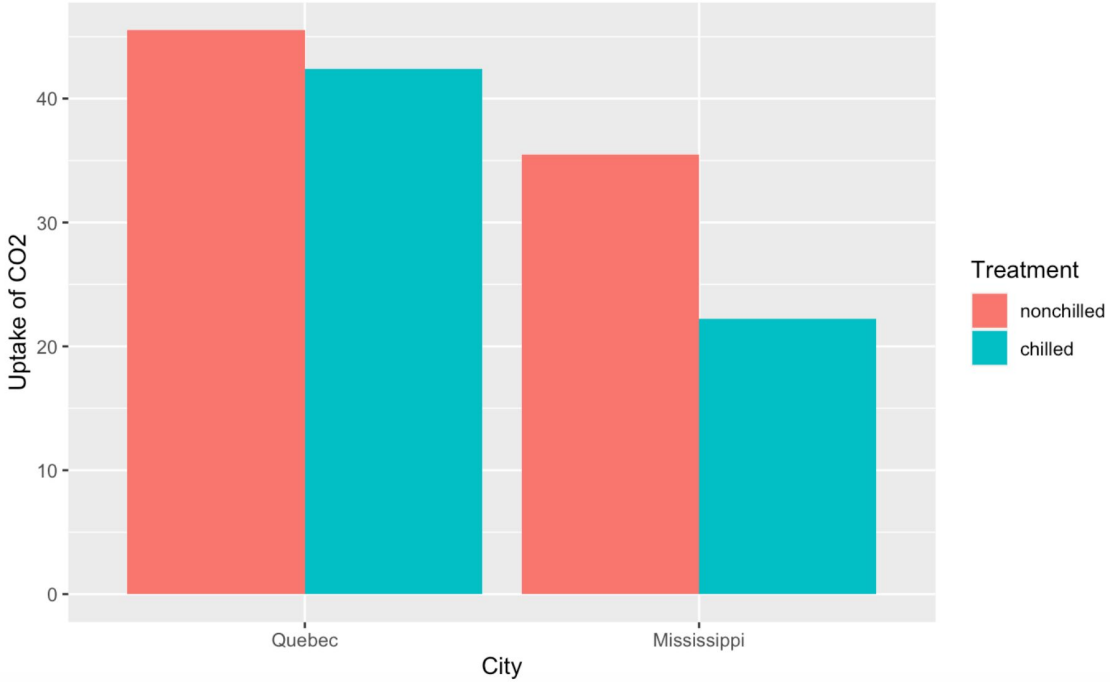


Stacked Bar Plot

```
data(CO2)
library(ggplot2)
ggplot(CO2, aes(x=Type, y=uptake, fill=Treatment)) + geom_bar(stat="identity") + xlab("City") + ylab("Uptake of CO2")
```

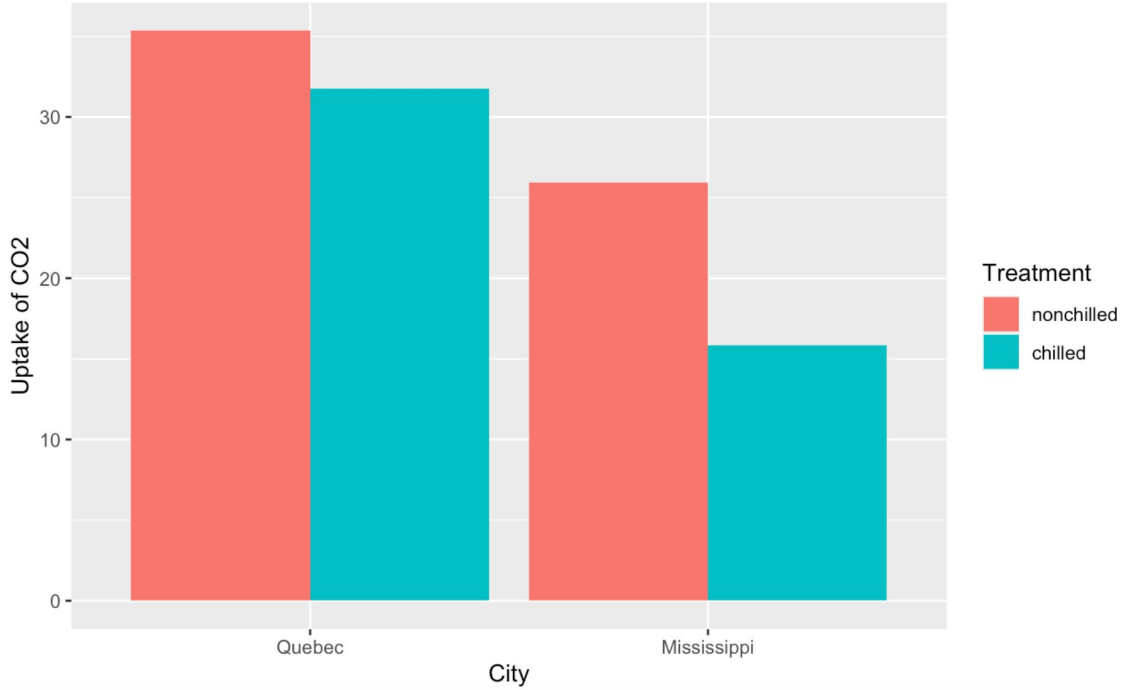
Grouped Bar Plot

```
data(CO2)
library(ggplot2)
ggplot(CO2, aes(x=Type, y=uptake, fill=Treatment)) + geom_bar(position="dodge", stat="identity") + xlab("City") +
 ylab("Uptake of CO2")
```



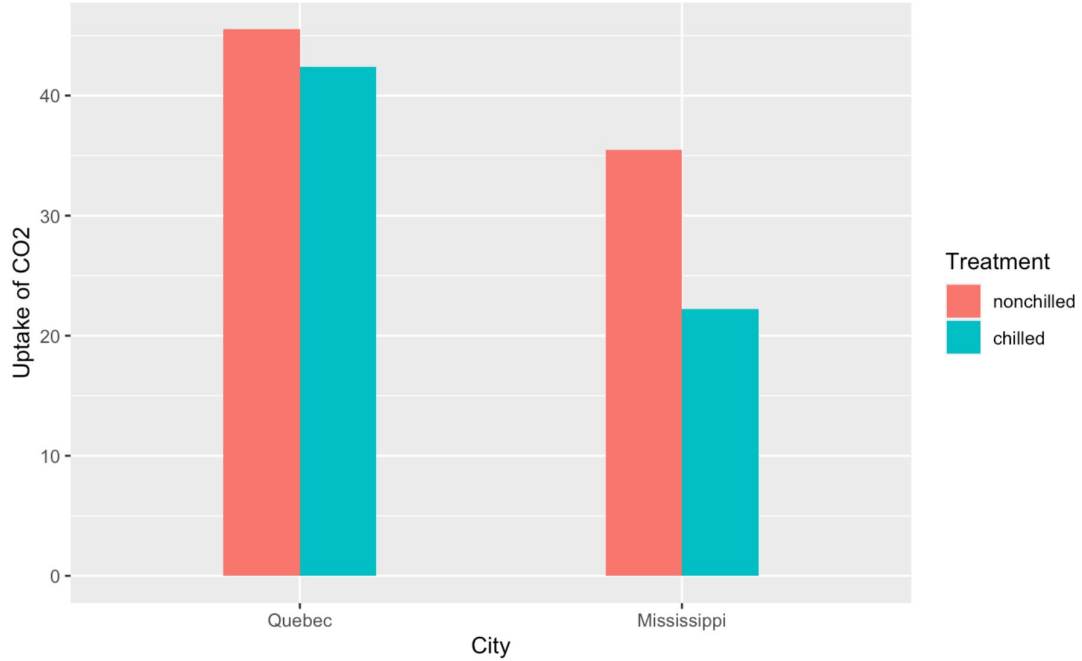
Mean

```
data(CO2)
library(ggplot2)
ggplot(CO2, aes(x=Type, y=uptake, fill=Treatment)) + geom_bar(position="dodge",stat="summary",fun.y="mean") + xla
b("City") + ylab("Uptake of CO2")
```



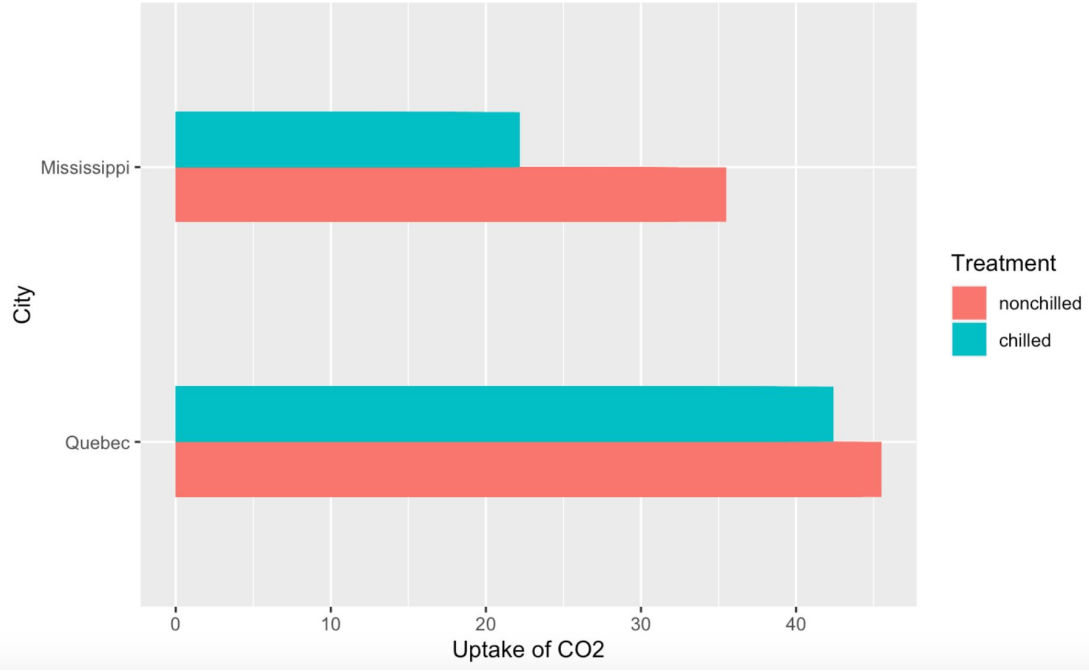
Width

```
data(CO2)
library(ggplot2)
ggplot(CO2, aes(x=Type, y=uptake, fill=Treatment)) + geom_bar(position="dodge", stat="identity", width=0.4) + xla
b("City") + ylab("Uptake of CO2")
```



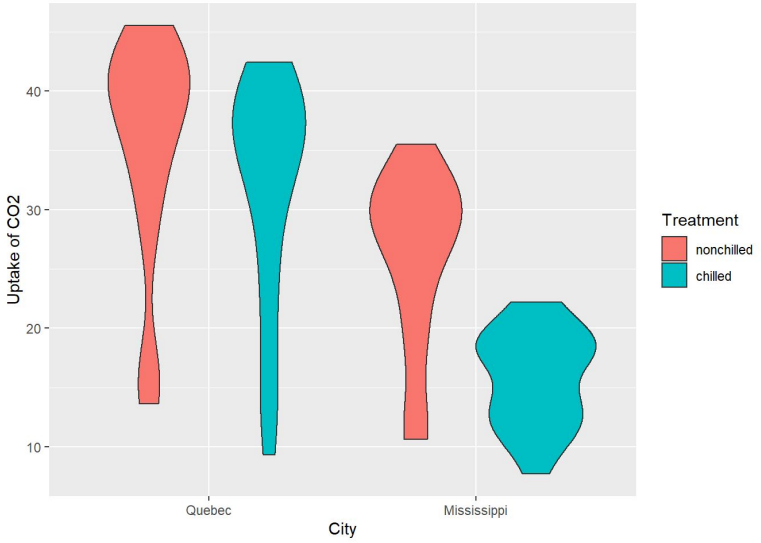
Horizontal Bar Plot

```
data(CO2)
library(ggplot2)
ggplot(CO2, aes(x=Type, y=uptake, fill=Treatment)) + geom_bar(position="dodge", stat="identity", width=0.4) + xla
b("City") + ylab("Uptake of CO2") + coord_flip()
```

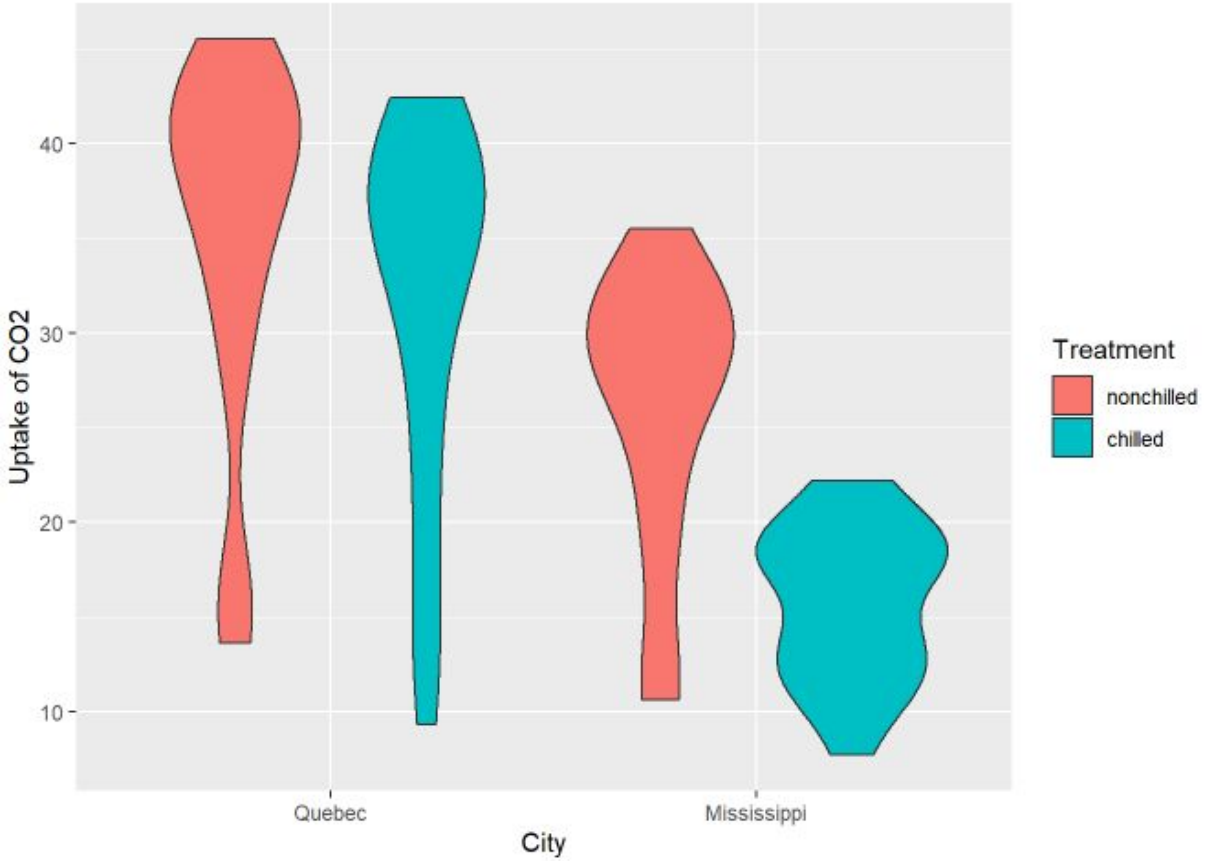


Violin Plots

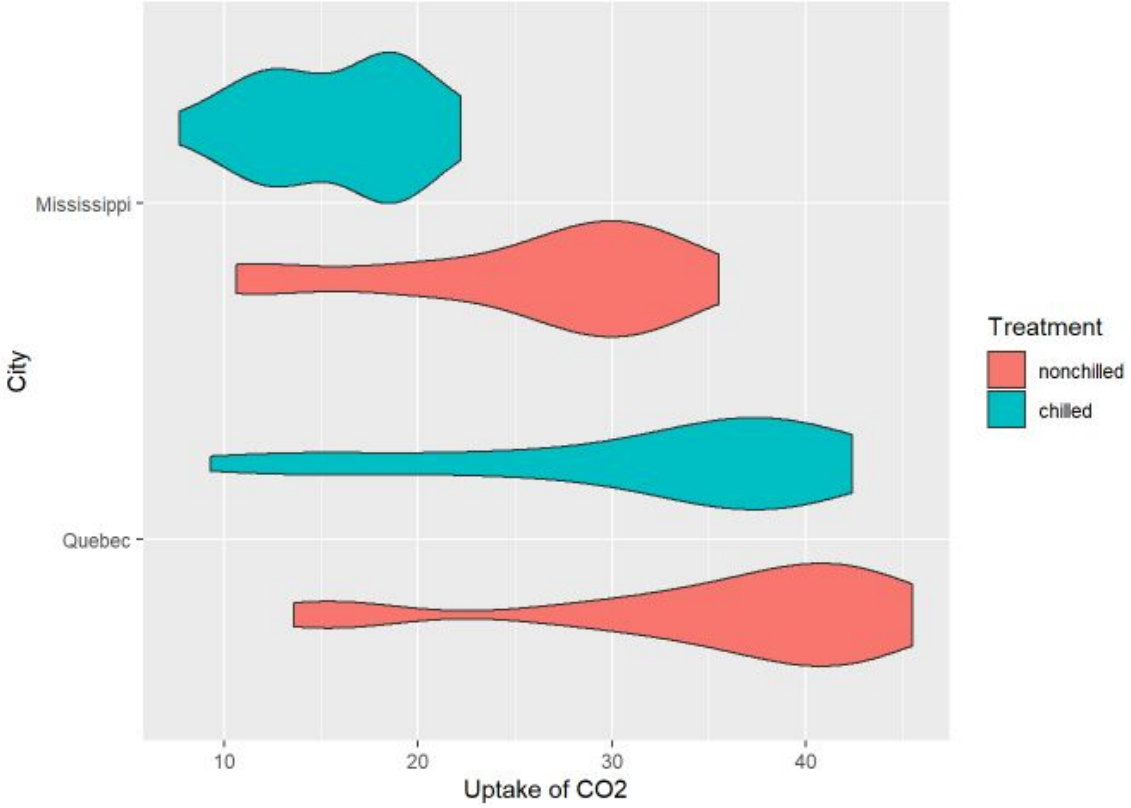
- Shows distribution of numeric data for multiple groups using density curves
- Usually these graphs are accompanied by a different chart type, like a box plot
- The next few slides show examples of the possible ways of producing bar graphs for the data set CO2 in R



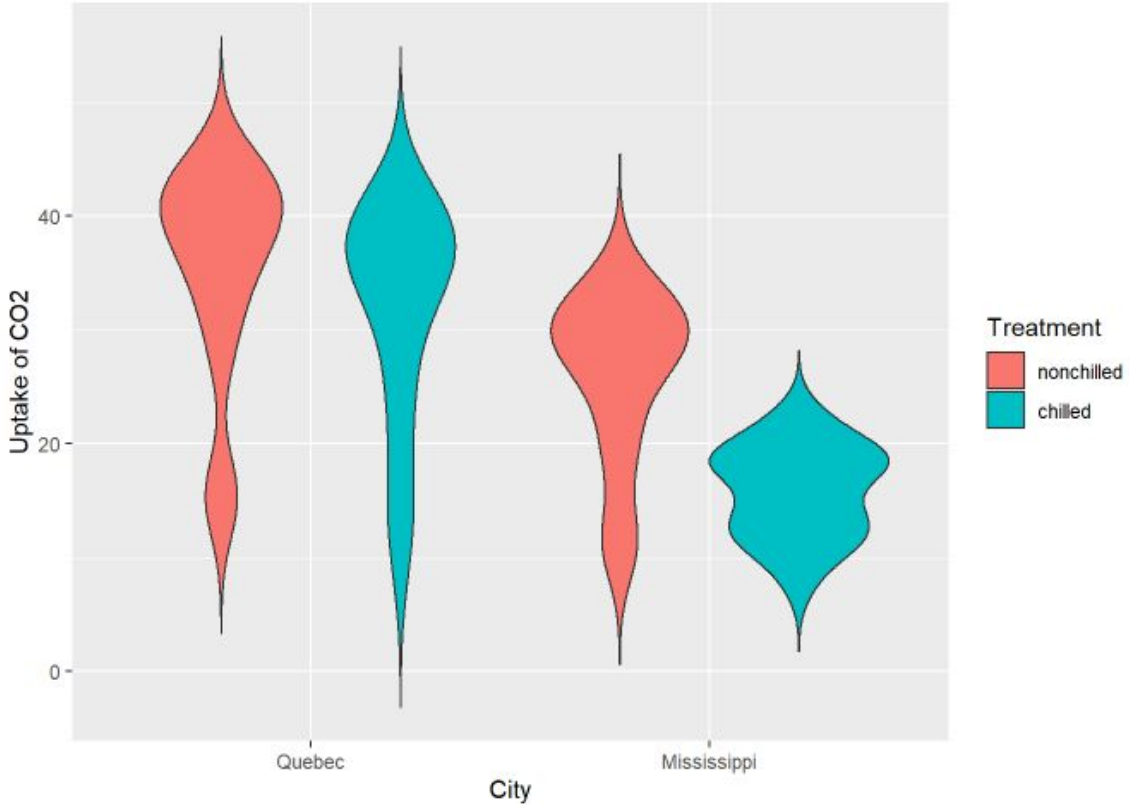
```
ggplot(CO2, aes(x= Type, y= uptake, fill = Treatment)) + geom_violin() + xlab("City") + ylab("Uptake of CO2")
```



```
ggplot(CO2, aes(x= Type, y= uptake, fill = Treatment)) + geom_violin() + xlab("City") + ylab("Uptake of CO2") + c
oord_flip()
```



```
ggplot(CO2, aes(x= Type, y= uptake, fill = Treatment)) + geom_violin(trim = FALSE) + xlab("City") + ylab("Uptake
 of CO2")
```



Resources

https://stackoverflow.com/questions/30183199/ggplot2-plot-mean-with-geom-bar

https://statisticsglobe.com/draw-grouped-barplot-in-r

http://www.sthda.com/english/wiki/ggplot2-barplots-quick-start-guide-r-software-and-data-visualization#basic-barplots

https://www.r-graph-gallery.com/218-basic-barplots-with-ggplot2.html

http://www.sthda.com/english/wiki/ggplot2-violin-plot-quick-start-guide-r-software-and-data-visualization