Biol 3350:
Data Analysis in Ecology

Data Exploration

Outline for today

- Debugging your code
- Digging into your data
- Today's assignment

Debugging your code: What is a warning?

Warnings indicate there might be a problem with your code

Debugging your code: What is a warning?

Warnings indicate there might be a problem with your code

```
> as.numeric(c("5", "6", "seven"))
[1]  5  6 NA
Warning message:
NAs introduced by coercion
```

Debugging your code: Errors!

Warnings indicate there might be a problem with your code

Errors mean the problem has forced R to stop executing your code

```
> as.numeric(c("5", "6", "seven"))
[1]  5  6 NA
Warning message:
NAs introduced by coercion
```

Debugging your code: Errors!

Warnings indicate there might be a problem with your code

Errors mean the problem has forced R to stop executing your code

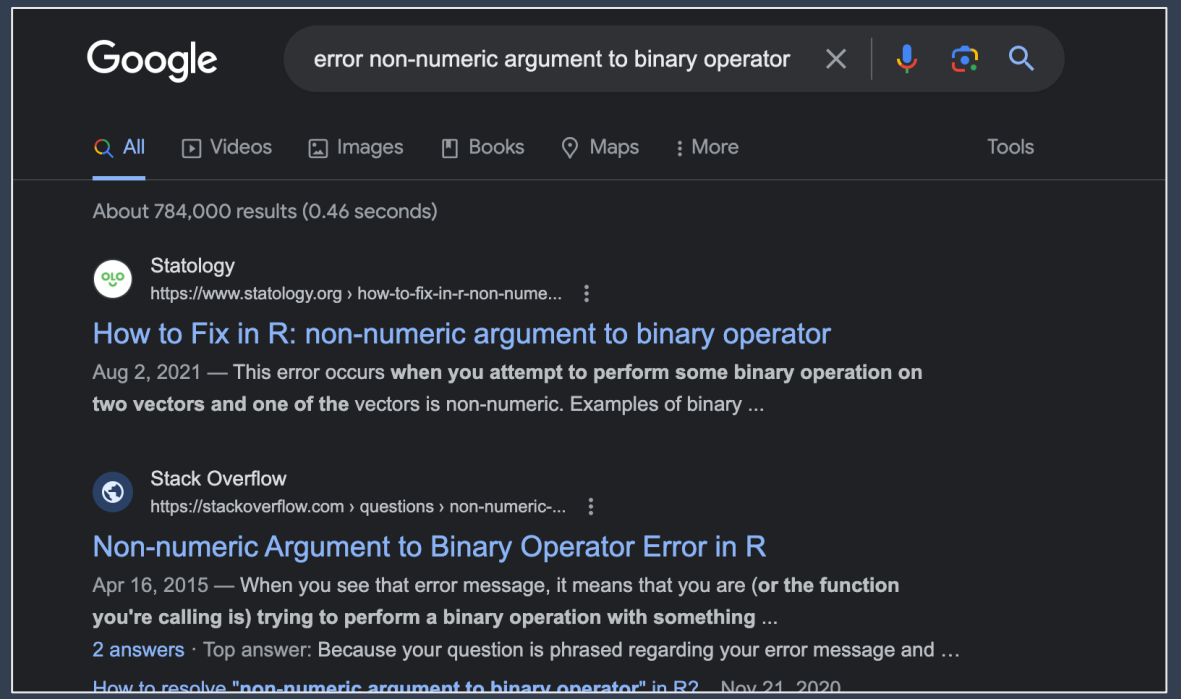
```
> as.numeric(c("5", "6", "seven"))
[1]  5  6 NA
Warning message:
NAs introduced by coercion
```

```
> "apple" + "banana" + 5
Error in "apple" + "banana" : non-numeric argument to binary operator
```

Debugging your code: Google is your best friend

Warnings indicate there might be a problem with your code

Errors mean the problem has forced R to stop executing your code



"Error non-numeric argument to binary operator in R"

error description from R

Digging into your data: Metadata

Metadata is the data that describes your data

Digging into your data: Metadata

Metadata is the data that describes your data

.txt file >>>>

README.txt

```
# The impact of dietary breadth on bumblebee colony fitness


We provide the following datasets, in order to follow all the statistical analyses of the
publication. These dataset appear, in order of appearance in the manuscript, along the R code also
submitted here:
###################################
Section 1: Plant diversity
###################################
Metabarcoding.txt:

Details for: Metabarcoding.txt
---------------------------------------

* Description: a tab-delimited txt file containing the metabarcoding output, thus identifying plant
taxa present in the pollen pellets as different OTU units, and the number of reads in each sample.
Pollen pellets are pooled for each collection date and site.
* Format(s): .txt

* Size(s): 12 KB

* Dimensions: 48 rows x 91 columns

* Variables:

  * ID: character variable providing unique identification code for an individual sample
  * Location: sampling site
  * Habitat: grassland or heathland,
  * Sampling: sampling occasion after placing hives in the field,
  * Columns 5 through 91: first row contains OTU names as assigned by the bioinformatics pipeline.
```

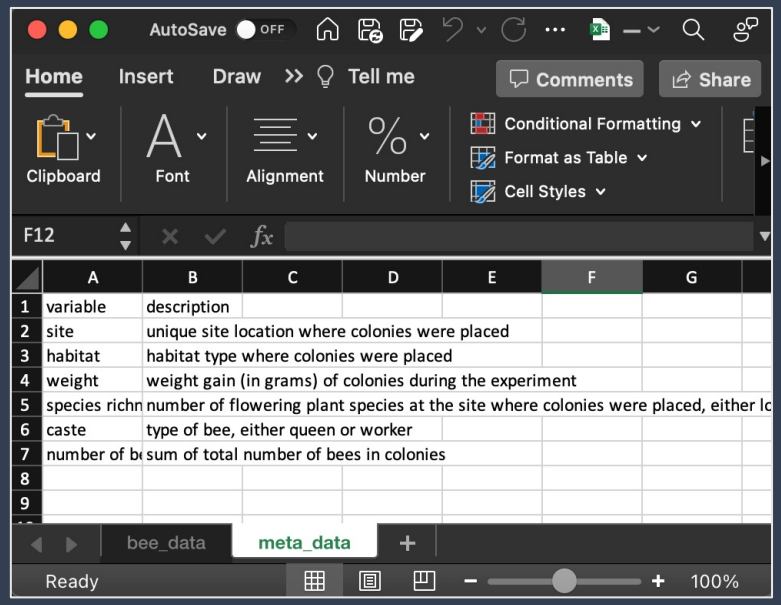
Digging into your data: Metadata

Metadata is the data that describes your data

.txt file >>>>

README.txt

```
# The impact of dietary breadth on bumblebee colony fitness


We provide the following datasets, in order to follow all the statistical analyses of the
publication. These dataset appear, in order of appearance in the manuscript, along the R code also
submitted here:
####################################
Section 1: Plant diversity
####################################
Metabarcoding.txt:

Details for: Metabarcoding.txt
----------------------------------------

* Description: a tab-delimited txt file containing the metabarcoding output, thus identifying plant
taxa present in the pollen pellets as different OTU units, and the number of reads in each sample.
Pollen pellets are pooled for each collection date and site.
* Format(s): .txt

* Size(s): 12 KB

* Dimensions: 48 rows x 91 columns

* Variables:

  * ID: character variable providing unique identification code for an individual sample
  * Location: sampling site
  * Habitat: grassland or heathland,
  * Sampling: sampling occasion after placing hives in the field,
  * Columns 5 through 91: first row contains OTU names as assigned by the bioinformatics pipeline.
```

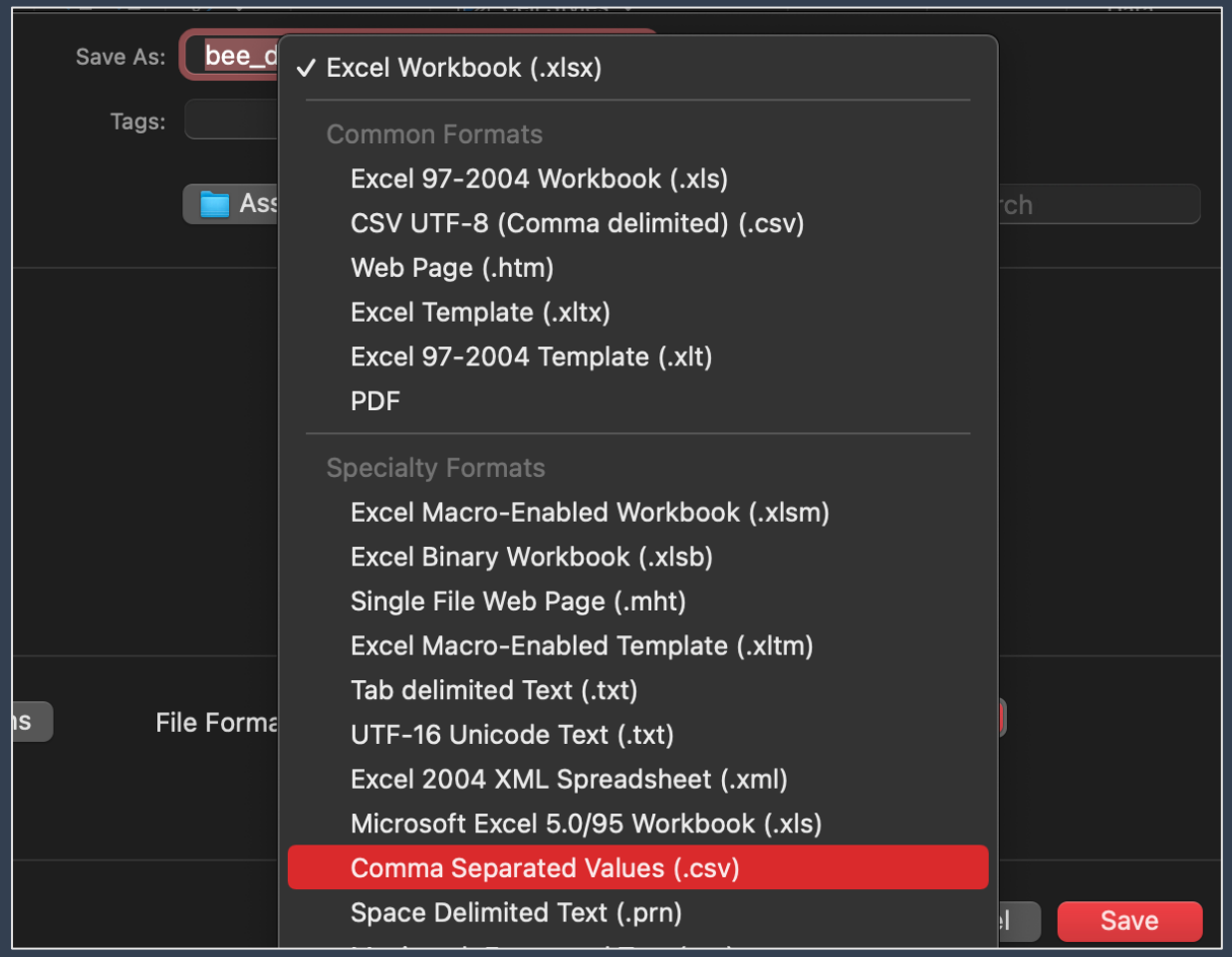


	A	B
1	variable	description
2	site	unique site location where colonies were placed
3	habitat	habitat type where colonies were placed
4	weight	weight gain (in grams) of colonies during the experiment
5	species richn	number of flowering plant species at the site where colonies were placed, either lo
6	caste	type of bee, either queen or worker
7	number of b	sum of total number of bees in colonies

<<<< .xlsx file

Digging into your data: Working with Excel files

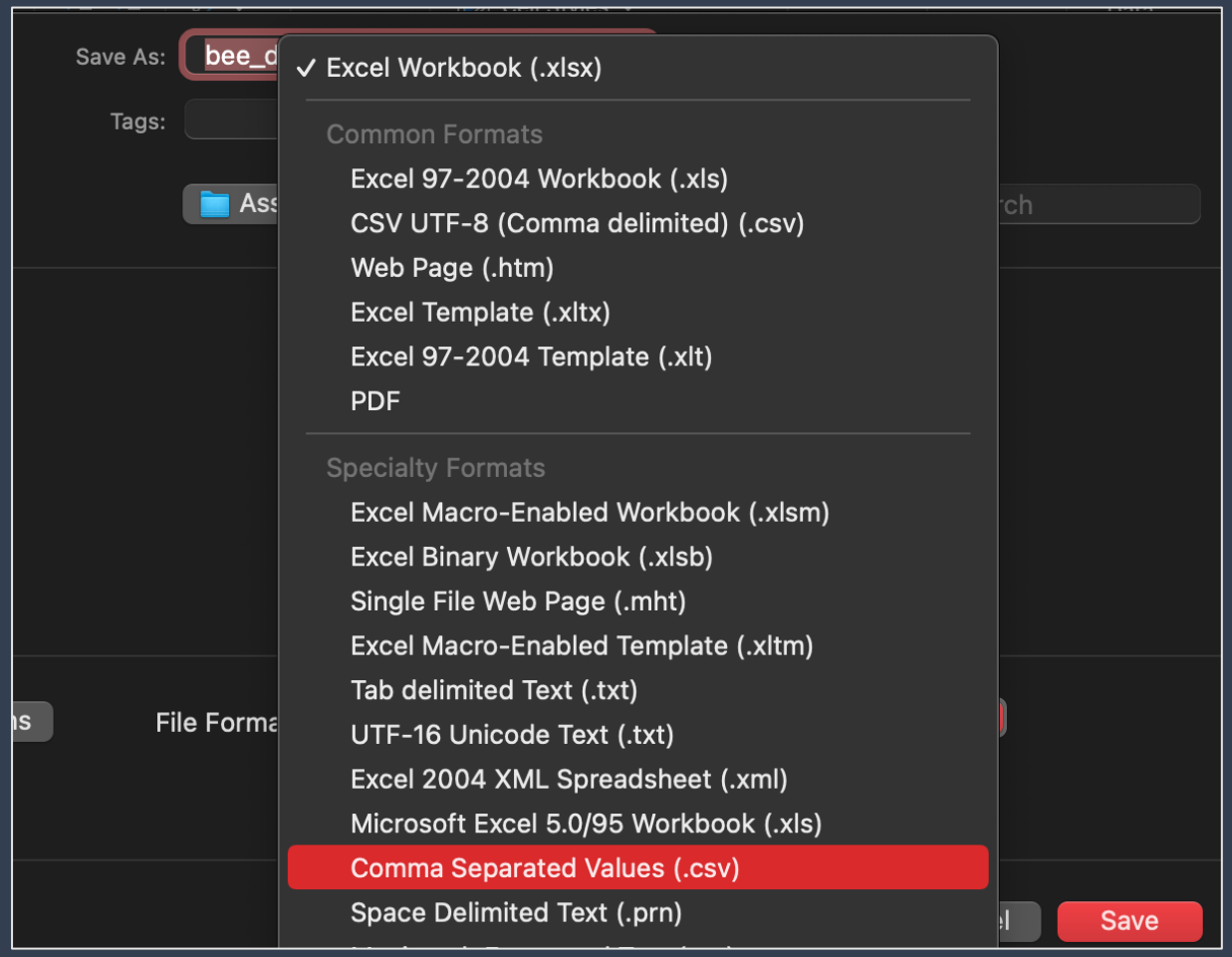
"Save as" your .xlsx file as a new .csv file



Digging into your data: Working with Excel files

"Save as" your .xlsx file as a new .csv file

When finished, move the .csv file to your working directory (or project folder)



Today's assignment: Learning objectives

- Recognize the key types of variables (in a statistical sense)
- Visually identify the structures of variables
- Visually evaluate relationships between variables using plots

Today's assignment: types of variables

Continuous variables indicate every possible value (in theory)

Today's assignment: types of variables

Continuous variables indicate every possible value (in theory)

Categorical variables can only be described by categories

Today's assignment: types of variables

Continuous variables indicate every value is possible (in theory)

Categorical variables can only be described by categories

Today's assignment: types of variables

Continuous variables indicate every value is possible (in theory)

Categorical variables can only be described by categories

numerics

characters >>

<< factors

Today's assignment: some familiar functions

rename() changes the names of individual variables

mutate() creates new columns with functions on existing variables

group_by() groups together like values of a variable

Today's assignment: some familiar functions

rename() changes the names of individual variables

mutate() creates new columns with functions on existing variables

group_by() groups together like values of a variable

summarize()/summarise()?

class()?

Today's assignment: new functions

filter() subsets a data frame by retaining all rows that satisfy a condition (e.g., filter(data.frame, variable == "value")

Today's assignment: new functions

filter() subsets a data frame by retaining all rows that satisfy a condition (e.g., filter(data.frame, variable == "value")

```
> ?filter
```

Use ***?*function name** to open documentation in your help files!

Today's assignment

New assignment format