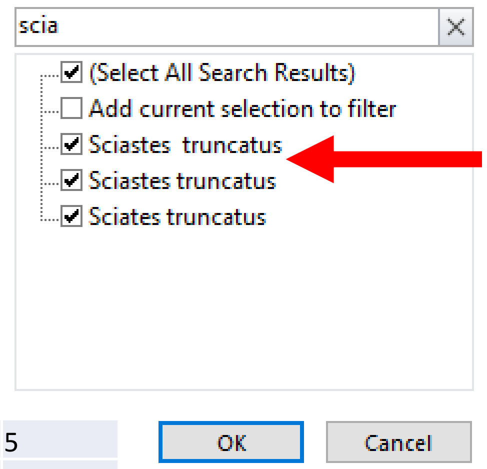
Data Management/Excel Hot-keys/Pivot Tables

This PDF contains a lot of useful excel things, some basics of data entry, and lots of text

Review: 3 Rules of Data

- No Blanks
- One Data Per Column
- No Repeat Rows (Duplications)

Rule 3.5: No Typos.


scia

☑ (Select All Search Results)
☐ Add current selection to filter
☑ Sciastes truncatus
☑ Sciastes truncatus
☑ Sciates truncatus

Typos (usually sneaky spaces)

harbinger of doom

Two data!!

SH-T21:0M-1	Linyphiidae	Sciastes	truncatus	Sciastes truncatus	Sciatrunca	M5	5
SH-T21:0M-1	Linyphiidae	Walckenaeria	karpinskii	Walckenaeria karpinskii	Walckarpin	M2	2
SH-T21:0M-1	Linyphiidae	Walckenaeria	trichornis	Walckenaeria trichornis	Walctricho	M2	2
SH-T21:0M-1	Thomisidae	Xysticus	britcheri	Xysticus britcheri	Xystbritch	M3	3
SH-T21:0M-3	Linyphiidae	Sisicottus	montanus	Sisicottus montanus	Sisimontan	F	1
SH-T21:0M-3	Amaurobidae	Amaurobidae	juvenile	Amaurobidae juvenile	Amaujuveni	NA	
SH-T21:0M-3	Linyphiidae	Sciastes	truncatus	Sciastes truncatus	Sciatrunca	F	1
SH-T21:0M-3	Linyphiidae	Linyphiidae	juvenile	Linyphiidae juvenile	Linyjuveni	NA	3

Blanks = Bad

Duplication = bad

Should have been entered as P. dist F 6, P moe 1

C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	distincta	F		1
C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	distincta	F		1
C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	distincta	F		1
C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	distincta	F		1
C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	moesta	F		1
C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	distincta	F		1
C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	Distincta	F		1

TrapID on every row can also lead to problems later.

Review: Title things efficiently

Absolutely nightmares to call in R, Use _ **not spaces** (also make names short)

If you are working with long name data, consider creating a column for short-hands

Sample ID	PFTID	Interior or Edge	Field Number	Deployment Date	Retrieval Date	Family	Genus	Species	SpeciesID	sex	count	Determiner	comments

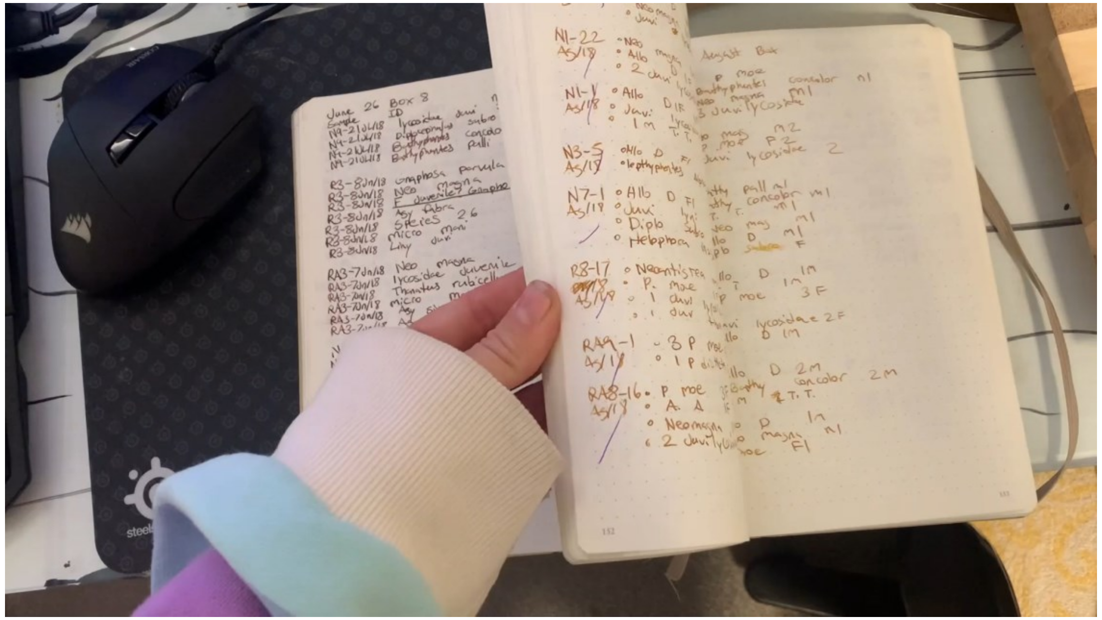
Easy to call in R.

Sample ID	PFTID	Habitat	Field_No	Deploy	Retrieve	Family	Genus	Species	SpeciesID	sex	count	Determiner	comments

(R likes underscores better than dashes, it treats – as minus, so preferentially use _ in titles)

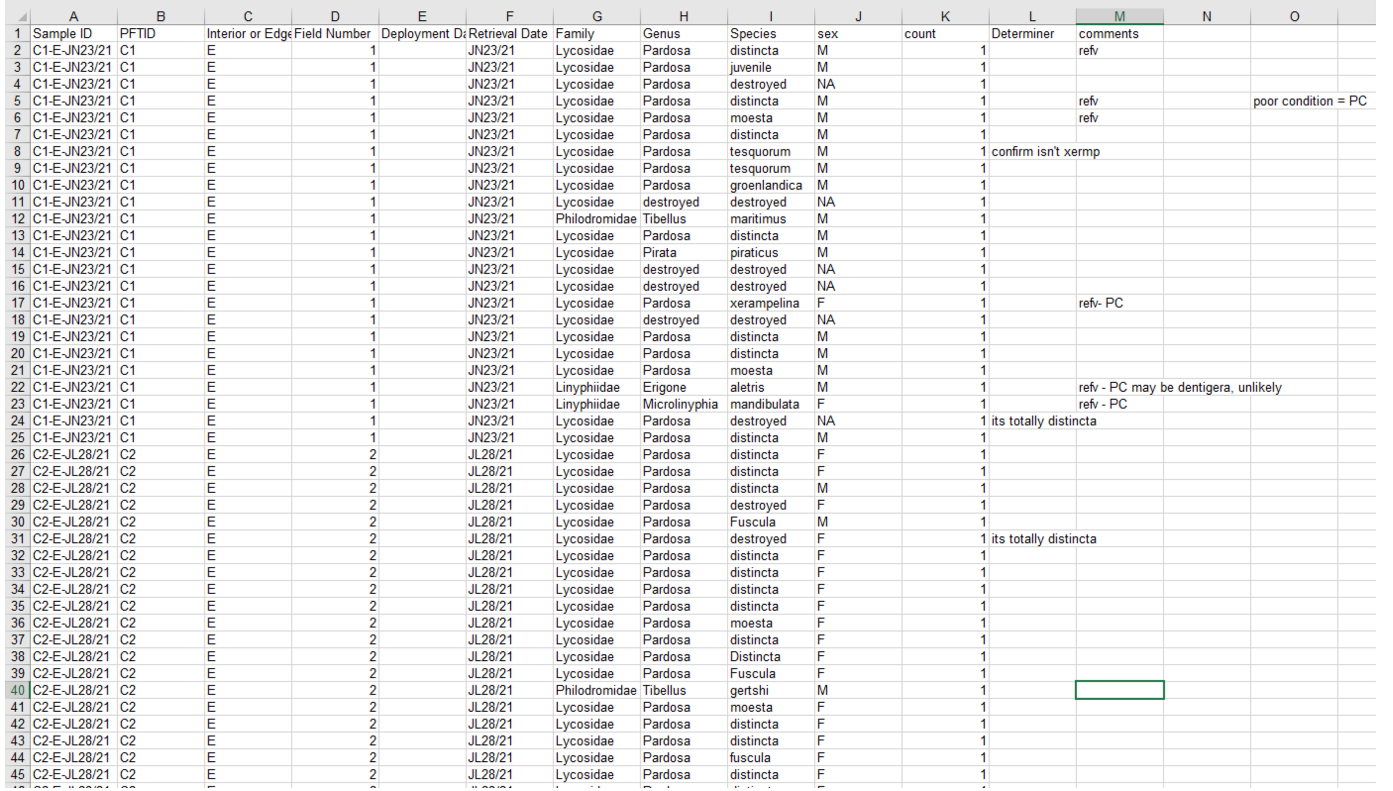
Bonus Note: Don’t put numbers at the start of your titles, R doesn’t like it

Initial Entry of Data



Write it down and transfer it: This adds anywhere from a few days to a few weeks of extra working time to your project. Additionally, can lead to errors if you have someone else do it for you and they are unfamiliar with your coding system.

Make decisions appropriate for your data collection method.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Sample ID	PFTID	Interior or Edge	Field Number	Deployment Da	Retrieval Date	Family	Genus	Species	sex	count	Determiner	comments		
2	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	distincta	M	1		refv		
3	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	juvenile	M	1				
4	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	destroyed	NA	1				
5	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	distincta	M	1		refv		poor condition = PC
6	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	moesta	M	1		refv		
7	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	distincta	M	1				
8	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	tesquorum	M	1	confirm isn't xermp			
9	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	tesquorum	M	1				
10	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	groenlandica	M	1				
11	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	destroyed	destroyed	NA	1				
12	C1-E-JN23/21	C1	E	1		JN23/21	Philodromidae	Tibellus	maritimus	M	1				
13	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	distincta	M	1				
14	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pirata	piraticus	M	1				
15	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	destroyed	destroyed	NA	1				
16	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	destroyed	destroyed	NA	1				
17	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	xerampelina	F	1		refv- PC		
18	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	destroyed	destroyed	NA	1				
19	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	distincta	M	1				
20	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	distincta	M	1				
21	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	moesta	M	1				
22	C1-E-JN23/21	C1	E	1		JN23/21	Linyphiidae	Erigone	aletris	M	1		refv - PC may be dentigera, unlikely		
23	C1-E-JN23/21	C1	E	1		JN23/21	Linyphiidae	Microlinyphia	mandibulata	F	1		refv - PC		
24	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	destroyed	NA	1	its totally distincta			
25	C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	distincta	M	1				
26	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	distincta	F	1				
27	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	distincta	F	1				
28	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	distincta	M	1				
29	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	destroyed	F	1				
30	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	Fuscula	M	1				
31	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	destroyed	F	1	its totally distincta			
32	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	distincta	F	1				
33	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	distincta	F	1				
34	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	distincta	F	1				
35	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	distincta	F	1				
36	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	moesta	F	1				
37	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	distincta	F	1				
38	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	Distincta	F	1				
39	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	Fuscula	F	1				
40	C2-E-JL28/21	C2	E	2		JL28/21	Philodromidae	Tibellus	gertshi	M	1				
41	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	moesta	F	1				
42	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	distincta	F	1				
43	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	distincta	F	1				
44	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	fuscula	F	1				
45	C2-E-JL28/21	C2	E	2		JL28/21	Lycosidae	Pardosa	distincta	F	1				

With **good version management** and **including comments,** direct data entry is *often* more efficient and has less potential for error as there is only a single stage of information transfer.

This is potentially a controversial opinion.

Super basic hotkey reminder:

CTRL + A is select only cells with data

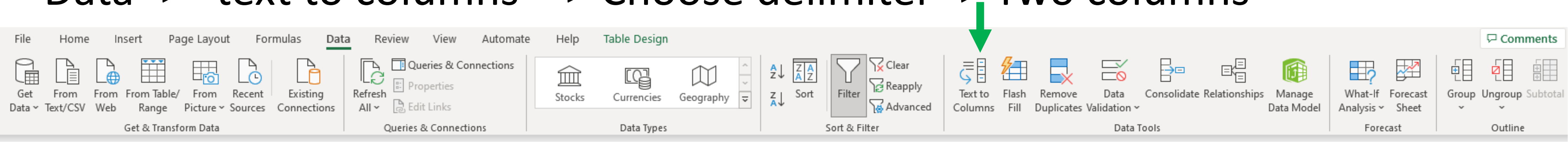
- Useful when you have 8000+ lines of data!
- Frequently forgotten hotkey!

<u>If you do the corner arrow select method your PC will crash when you make a pivot table.</u>

The rest of this PDF is text and excel boxes so here's a photo of my cats.

Excel Tools: Data Cleanup Strategies

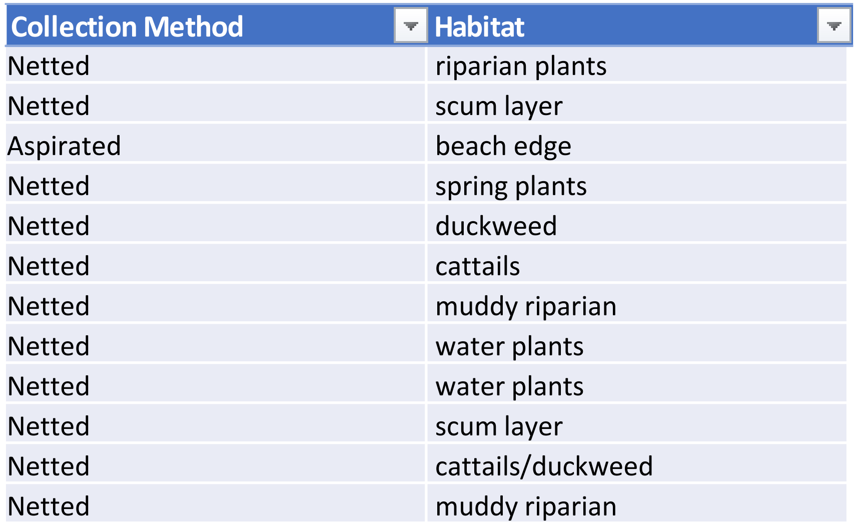
- **Column breakup delimiters** (Can fix two data!! One column! Problem)
- Data -> “text to columns” -> Choose delimiter -> Two columns



Two data! Ahh!

Collection Method
Netted, riparian plants
Netted, scum layer
Aspirated, beach edge
Netted, spring plants
Netted, duckweed
Netted, cattails
Netted, muddy riparian
Netted, water plants
Netted, water plants
Netted, scum layer
Netted, cattails/duckweed
Netted, muddy riparian

One data.



Collection Method	Habitat
Netted	riparian plants
Netted	scum layer
Aspirated	beach edge
Netted	spring plants
Netted	duckweed
Netted	cattails
Netted	muddy riparian
Netted	water plants
Netted	water plants
Netted	scum layer
Netted	cattails/duckweed
Netted	muddy riparian

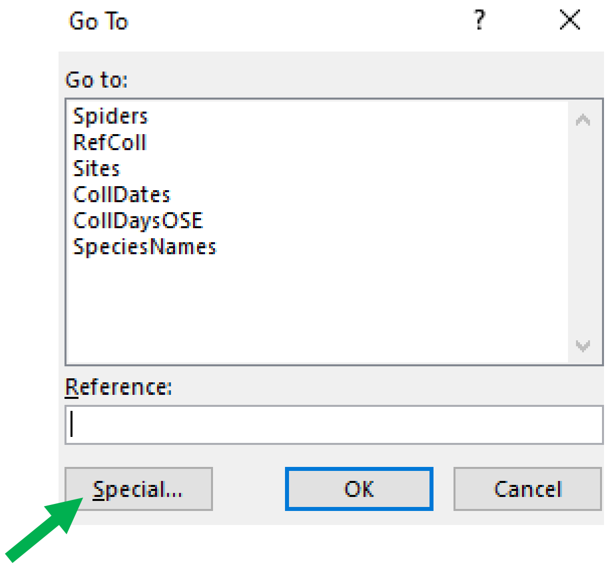
Important:

Before you use the delimiter tool make a blank column on the right side of the data you are breaking up. It will tell you there is already data in that column even if it is empty!

Excel Hotkey: Data Cleanup Strategies

- **Excel Blanks Fill**

1. CTRL + G opens “Go To”
2. Select Special
3. Select Blanks
4. type “=“ in cell box
5. Click Up Arrow
6. Click CTRL + Enter

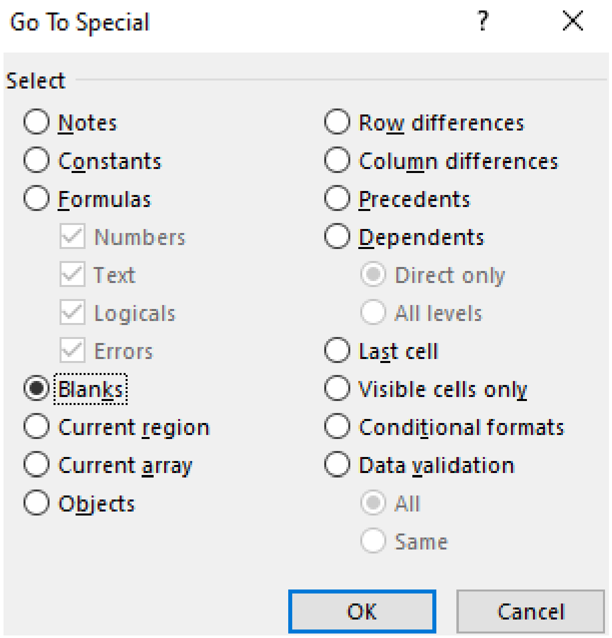


Go To ? ×

Go to:

Spiders
RefColl
Sites
CollDates
CollDaysOSE
SpeciesNames

Reference:



Go To Special ? ×

Select

○ Notes
○ Constants
○ Formulas
☑ Numbers
☑ Text
☑ Logicals
☑ Errors
● Blanks
○ Current region
○ Current array
○ Objects
○ Row differences
○ Column differences
○ Precedents
○ Dependents
● Direct only
○ All levels
○ Last cell
○ Visible cells only
○ Conditional formats
○ Data validation
● All
○ Same

This will fill all the blanks you’ve selected with the data directly above it.

Good for: When you have multiple collection site IDs and want to utilize a duplicate entry checker. Trim Function. More on these later!

Data Setup Strategy

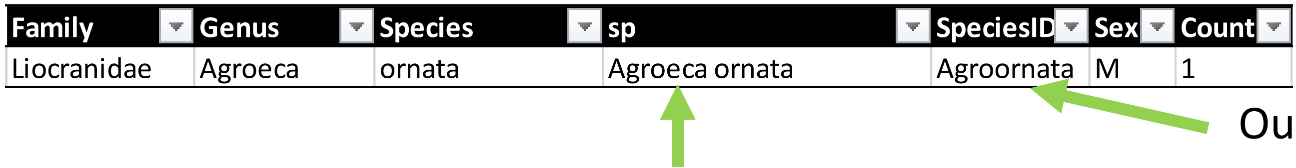
The R-Friendly Column:

When creating your data file, create an ID column. In my case, it is SpeciesID.

Convert your data into a table with titles

In the ID Column, put a code similar to this that is appropriate for your data

=LEFT([@Genus],4)&LEFT([@Species],6)



Family	Genus	Species	sp	SpeciesID	Sex	Count
Liocranidae	Agroeca	ornata	Agroeca ornata	Agroornata	M	1

Output of R-Friendly column

This one is =[@Genus]&" "&[@Species]

R-Friendly because:

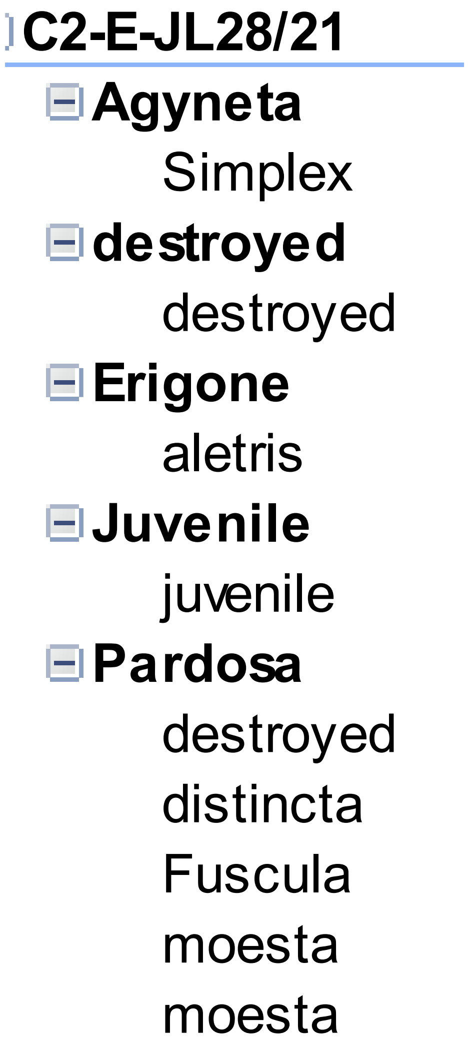
1) Easy for you to $ call in a data matrix (it’s not very long compared to many species names)
2) Automatically removes the space in the species name
3) Short enough that if you have multiple on a PCA your GGplot isn’t an unreadable word blob.

Neat Note:

if you add this column later on, paste the code into the first row of your **new column within** your **data table**, it will auto fill to the bottom!

The R-Friendly Column (R-FC): Same Sample Data:

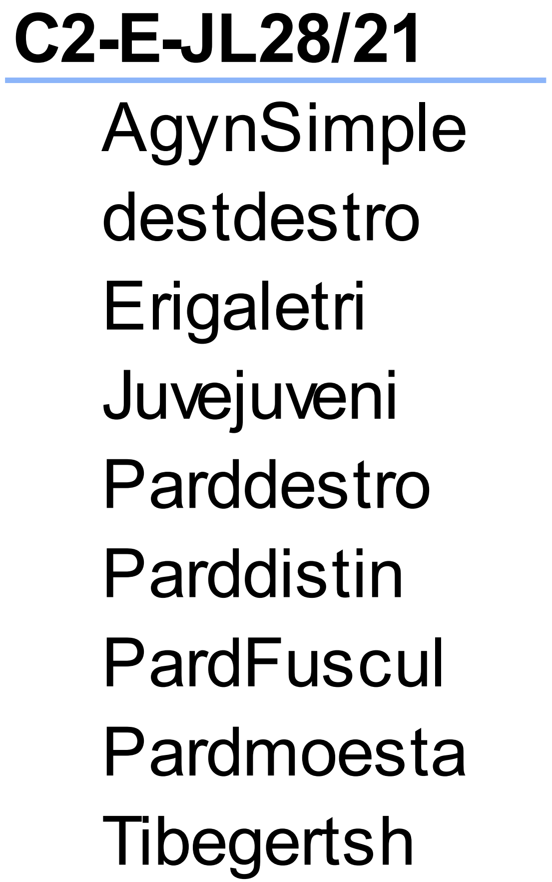
Without R-FC



C2-E-JL28/21

⊟ **Agyneta**
    Simplex
⊟ **destroyed**
    destroyed
⊟ **Erigone**
    aletris
⊟ **Juvenile**
    juvenile
⊟ **Pardosa**
    destroyed
    distincta
    Fuscula
    moesta
    moesta

With R-FC

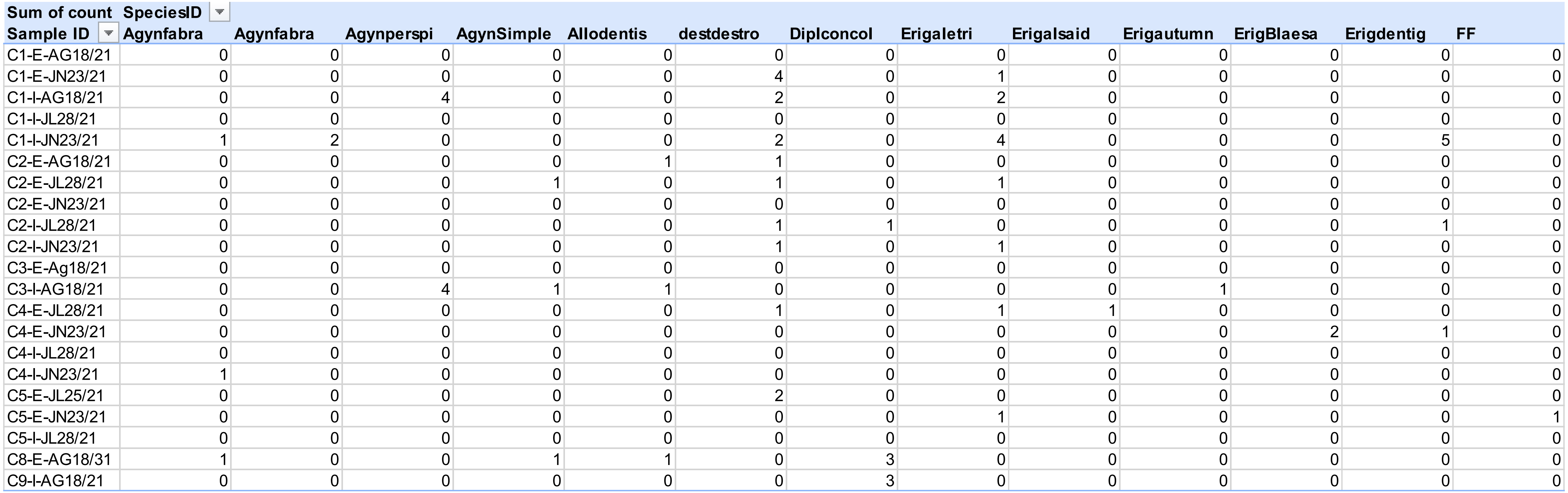


C2-E-JL28/21

AgynSimple
destdestro
Erigaletri
Juvejuveni
Parddestro
Parddistin
PardFuscul
Pardmoesta
Tibegertsh

The R-Friendly Column: R Ready Output

This was generated with a pivot table

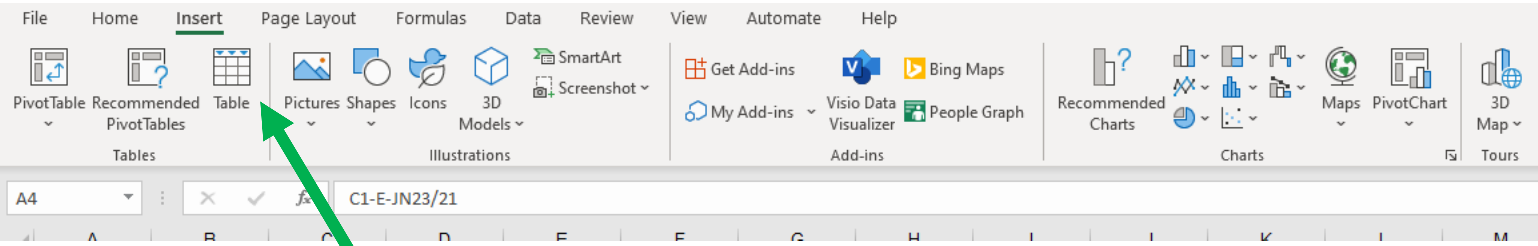


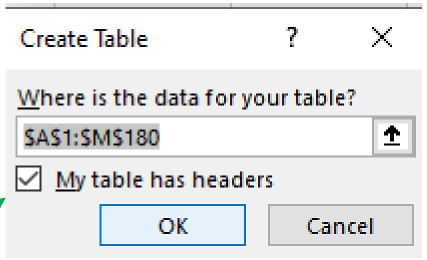
Sum of count	**SpeciesID** ▾												
Sample ID ▾	**Agynfabra**	**Agynfabra**	**Agynperspi**	**AgynSimple**	**Allodentis**	**destdestro**	**Diplconcol**	**Erigaletri**	**Erigalsaid**	**Erigautumn**	**ErigBlaesa**	**Erigdentig**	**FF**
C1-E-AG18/21	0	0	0	0	0	0	0	0	0	0	0	0	0
C1-E-JN23/21	0	0	0	0	0	4	0	1	0	0	0	0	0
C1-I-AG18/21	0	0	4	0	0	2	0	2	0	0	0	0	0
C1-I-JL28/21	0	0	0	0	0	0	0	0	0	0	0	0	0
C1-I-JN23/21	1	2	0	0	0	2	0	4	0	0	0	5	0
C2-E-AG18/21	0	0	0	0	1	1	0	0	0	0	0	0	0
C2-E-JL28/21	0	0	0	1	0	1	0	1	0	0	0	0	0
C2-E-JN23/21	0	0	0	0	0	0	0	0	0	0	0	0	0
C2-I-JL28/21	0	0	0	0	0	1	1	0	0	0	0	1	0
C2-I-JN23/21	0	0	0	0	0	1	0	1	0	0	0	0	0
C3-E-Ag18/21	0	0	0	0	0	0	0	0	0	0	0	0	0
C3-I-AG18/21	0	0	4	1	1	0	0	0	0	1	0	0	0
C4-E-JL28/21	0	0	0	0	0	1	0	1	1	0	0	0	0
C4-E-JN23/21	0	0	0	0	0	0	0	0	0	0	2	1	0
C4-I-JL28/21	0	0	0	0	0	0	0	0	0	0	0	0	0
C4-I-JN23/21	1	0	0	0	0	0	0	0	0	0	0	0	0
C5-E-JL25/21	0	0	0	0	0	2	0	0	0	0	0	0	0
C5-E-JN23/21	0	0	0	0	0	0	0	1	0	0	0	0	1
C5-I-JL28/21	0	0	0	0	0	0	0	0	0	0	0	0	0
C8-E-AG18/31	1	0	0	1	1	0	3	0	0	0	0	0	0
C9-I-AG18/21	0	0	0	0	0	0	3	0	0	0	0	0	0

Pivot tables: Make a table first

Use tables because the little arrows let you filter things!

You can also sort things in A-Z order for the column of arrow selection.




Create Table

Where is the data for your table?

$A$1:$M$180

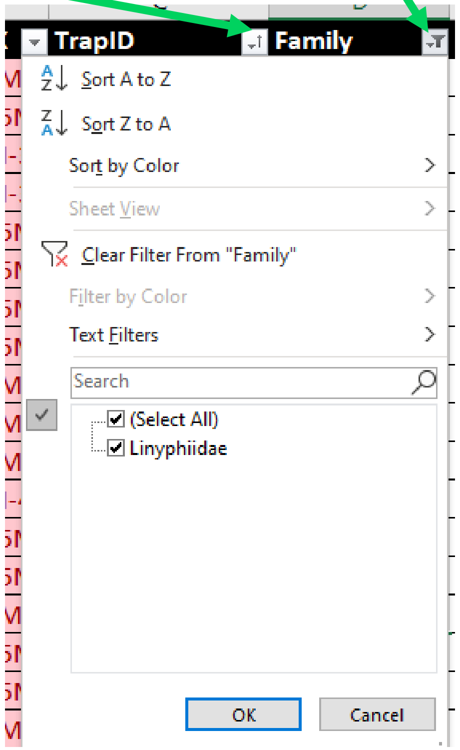
☑ My table has headers

Make sure this box is checked!

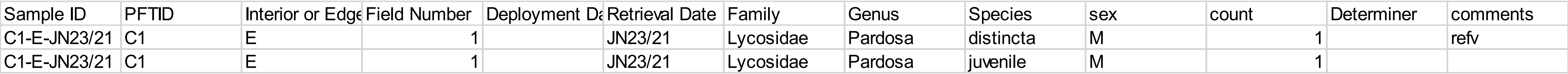
Note:

If you have colour-coded cells in your document, *before making the table*, excel may freeze when making a table.

This happens inconsistently, I don't know why. It will work eventually if it does freeze.



Before:

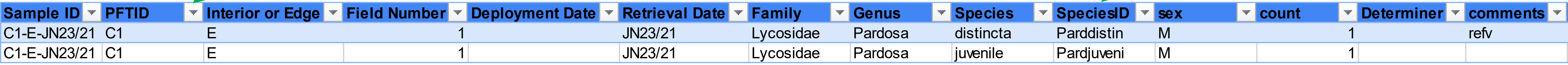


Sample ID	PFTID	Interior or Edge	Field Number	Deployment Da	Retrieval Date	Family	Genus	Species	sex	count	Determiner	comments
C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	distincta	M	1		refv
C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	juvenile	M	1		

After:

These arrows are your best friend with large data sets.

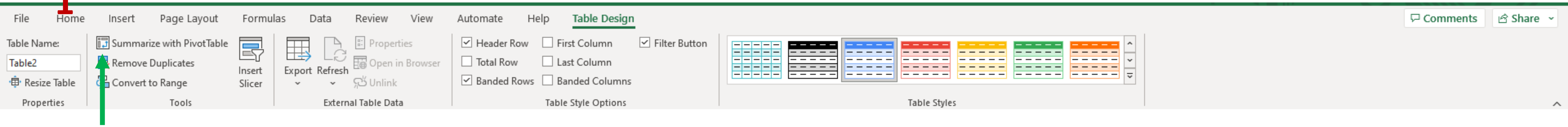
I added the R-Friendly column



Sample ID	PFTID	Interior or Edge	Field Number	Deployment Date	Retrieval Date	Family	Genus	Species	SpeciesID	sex	count	Determiner	comments
C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	distincta	Parddistin	M	1		refv
C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	juvenile	Pardjuveni	M	1		

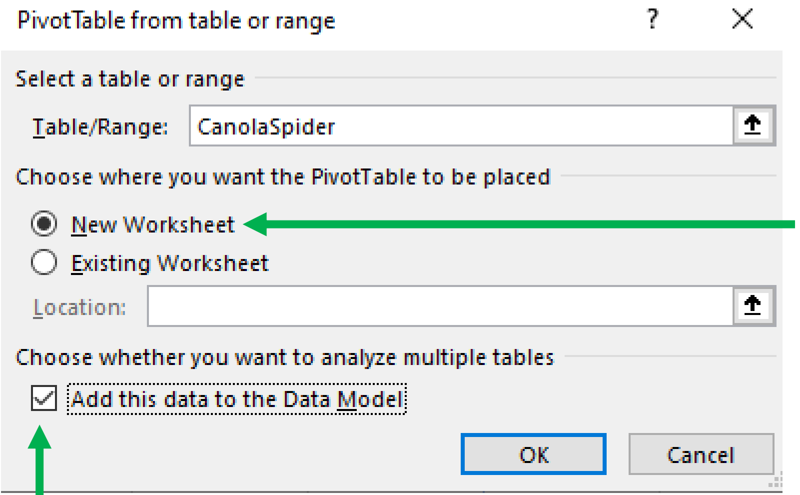
Pivot tables: Set Up Steps

1



Click this (have a cell within the data table clicked, make sure nothing is highlighted)
(name your table, its good practice, required for table linking)

2



Select a table or range

Table/Range: CanolaSpider

Choose where you want the PivotTable to be placed

● New Worksheet

○ Existing Worksheet

Location:

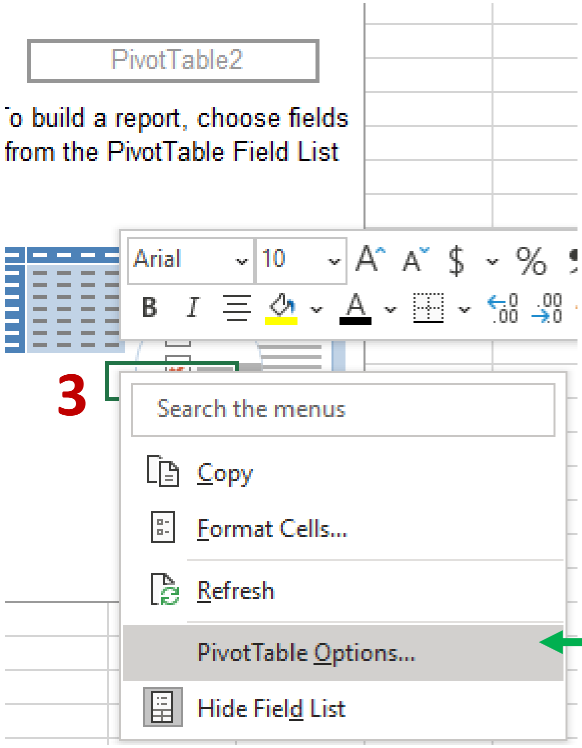
Choose whether you want to analyze multiple tables

☑ Add this data to the Data Model

Always add to new, adding to existing gets squirrely sometimes

Check this box, it allows you to do distinct counts/other math stuff I don't have time to cover

3



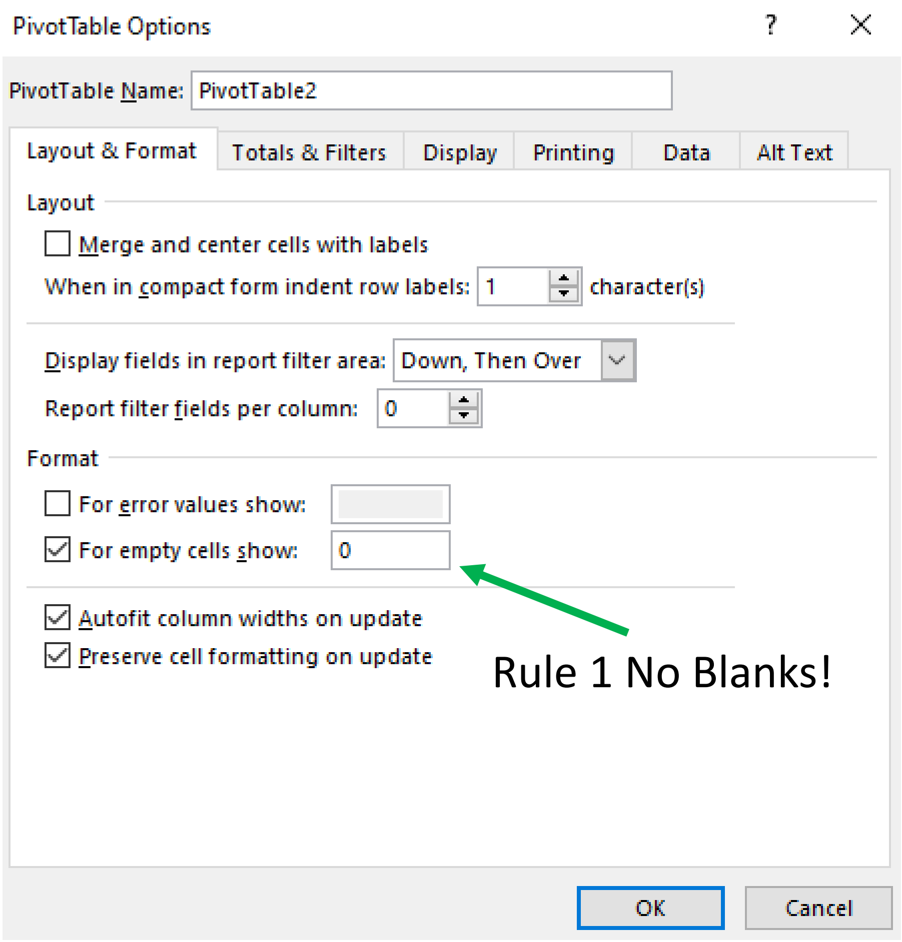
Right click, open Pivot table Options

4 ➡

Pivot tables: Options

There are two settings you want to change:

5



PivotTable Options

PivotTable Name: PivotTable2

Layout

☐ Merge and center cells with labels

When in compact form indent row labels: 1 character(s)

Display fields in report filter area: Down, Then Over

Report filter fields per column: 0

Format

☐ For error values show:

☑ For empty cells show: 0

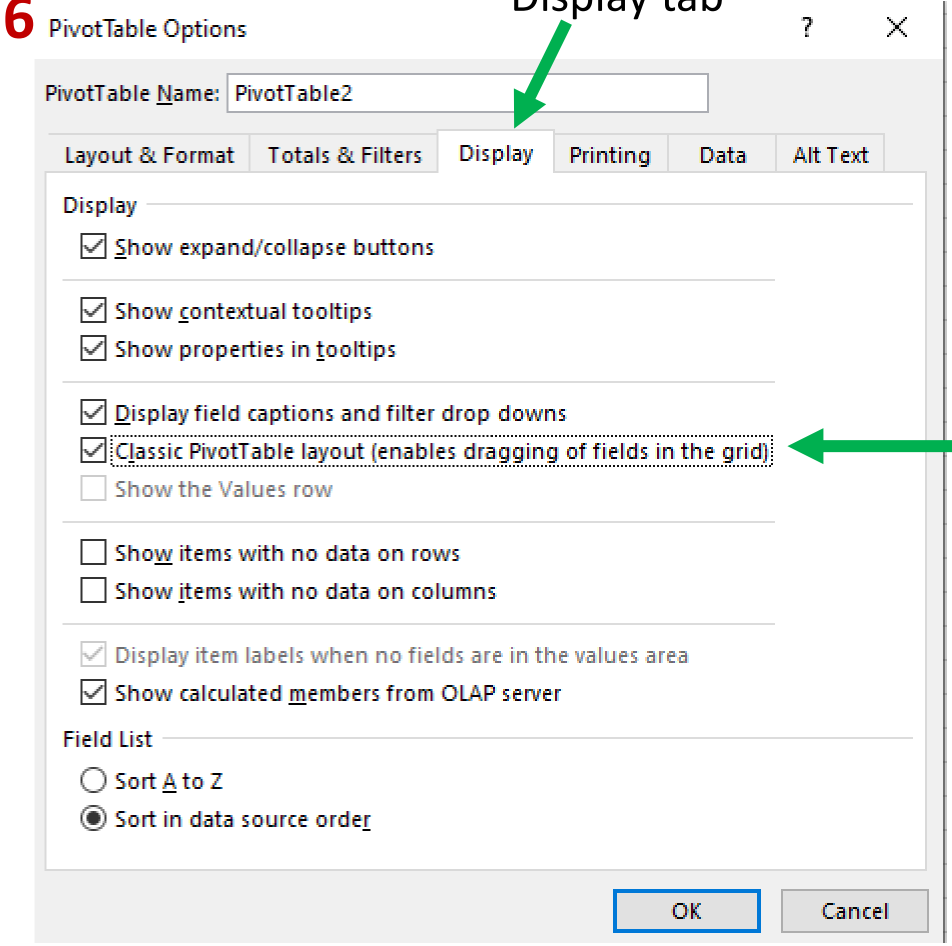
☑ Autofit column widths on update

☑ Preserve cell formatting on update

Rule 1 No Blanks!

6

Display tab



PivotTable Options

PivotTable Name: PivotTable2

Display

☑ Show expand/collapse buttons

☑ Show contextual tooltips

☑ Show properties in tooltips

☑ Display field captions and filter drop downs

☑ Classic PivotTable layout (enables dragging of fields in the grid)

☐ Show the Values row

☐ Show items with no data on rows

☐ Show items with no data on columns

☑ Display item labels when no fields are in the values area

☑ Show calculated members from OLAP server

Field List

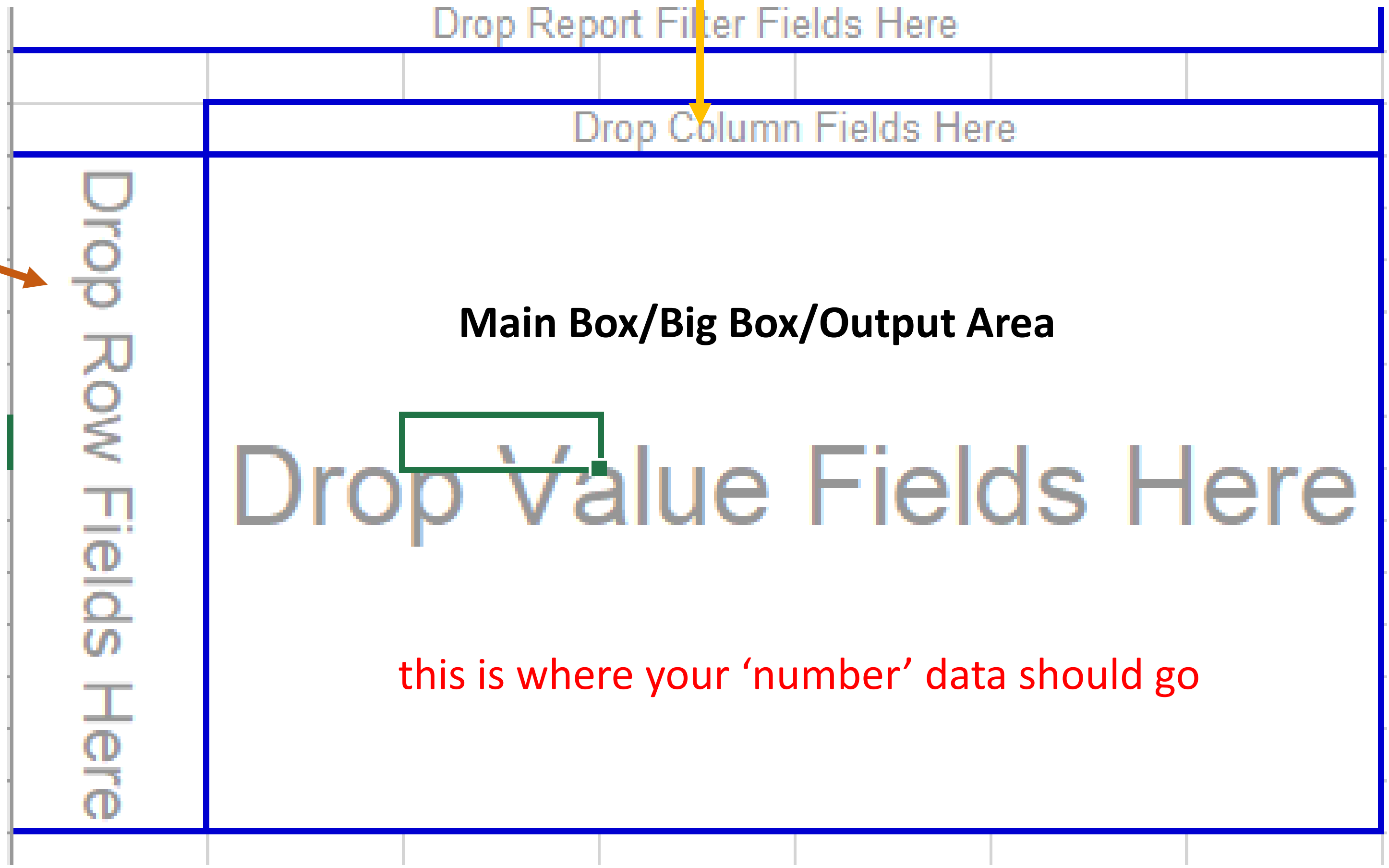
○ Sort A to Z

● Sort in data source order

Old table enables dragging, easier to use.

The Legendary Rectangle of (almost) unlimited power

Labelled diagram so you know what I am referring too throughout the rest of this presentation:



Horizontal Axis

Species names across the top

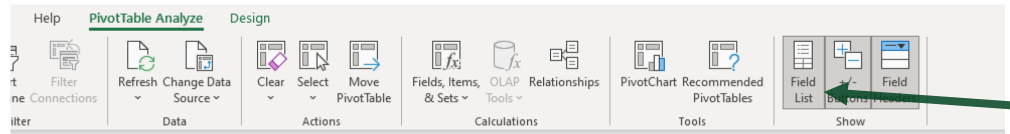
Vertical Axis

usually site information goes in vertical, but this is all dependent on the analysis tool you are using.

Main Box/Big Box/Output Area

this is where your ‘number’ data should go

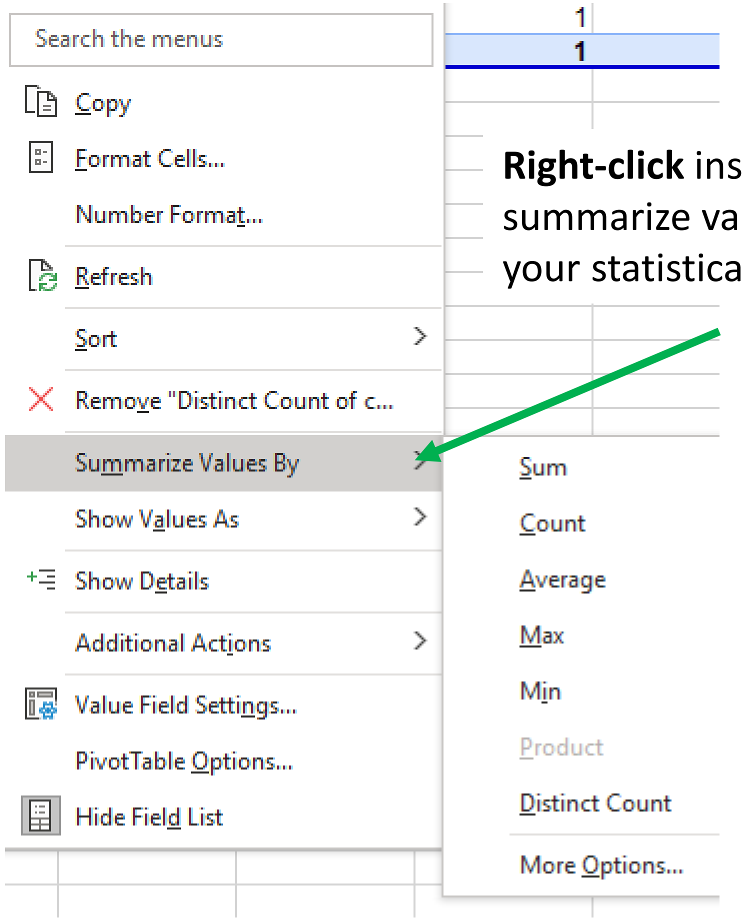
Data Setup



This button under **PivotTable analyze** turns the field list on and off, in case you accidentally lose your field list.

Choose the contents of PivotTable

1



Right-click inside **big box**, select summarize value that is appropriate for your statistical goals

2

Choose Axis Labels 1st

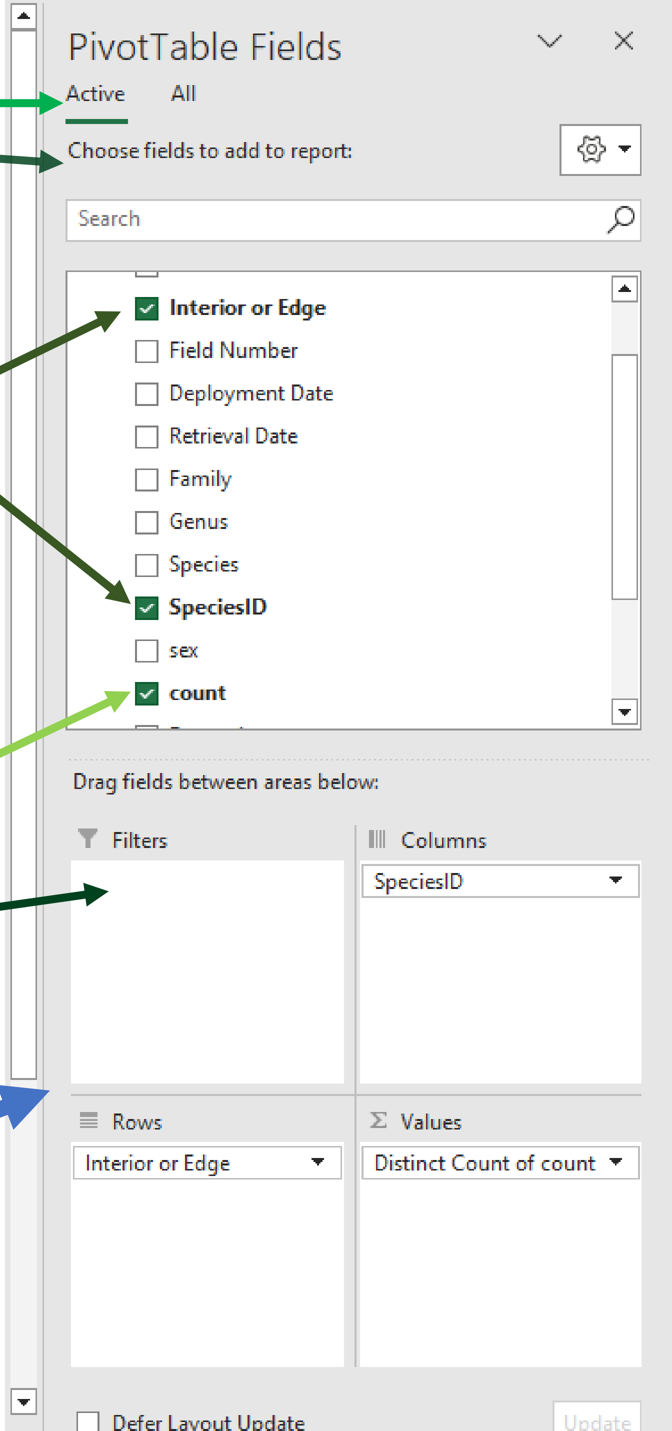
3

Choose input data 2nd

4

if you want to hide certain things in your data, use this field. I need to exclude juveniles from many analysis, so I filter those here.

You can also move stuff down here but I find working directly on the table easier to keep track of.



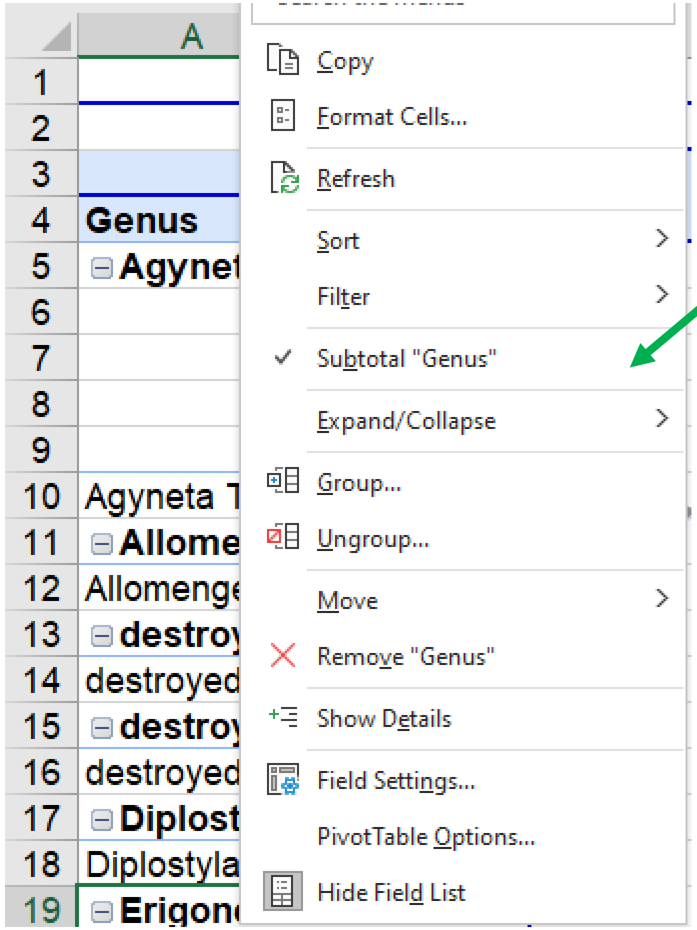
☑ Interior or Edge
☐ Field Number
☐ Deployment Date
☐ Retrieval Date
☐ Family
☐ Genus
☐ Species
☑ SpeciesID
☐ sex
☑ count

Columns: SpeciesID

Rows: Interior or Edge

Values: Distinct Count of count

Using Pivot Tables to Clean Up Data



Turn off subtotaling on your vertical axis. 99% of the time it is unhelpful.

	Genus	Species
4	Genus	Species
5	Agyneta	fabra
6		fabra
7		perspicua
8		perspicua
9		Simplex
10	Allomenge	dentisetis
11	destroyed	destroyed
12	destroyed	destroyed
13	Diplostyla	concolor
14	Erigone	aletris
15		alsaida
16		autumnalis
17		Blaesa
18		dentigera
19	Erigone	Blaesa
20	F	F
21	Grammonc	gentilis
22		gentilis
23	HOGNA	frondicola
24	Islandia	princeps
25	Islandia	princeps
26	Juvenile	juvenile
27	Micaria	juvenile
28		pulicaria
29	Microlinyp	mandibulata
30	Neoantiste	magna
31	Neottiura	bimaculata
32	Pardosa	damaged
33		destroyed
34		distincta
35		fuscla
36		Fuscula

This shows me:

I have invisible spaces all over my document.

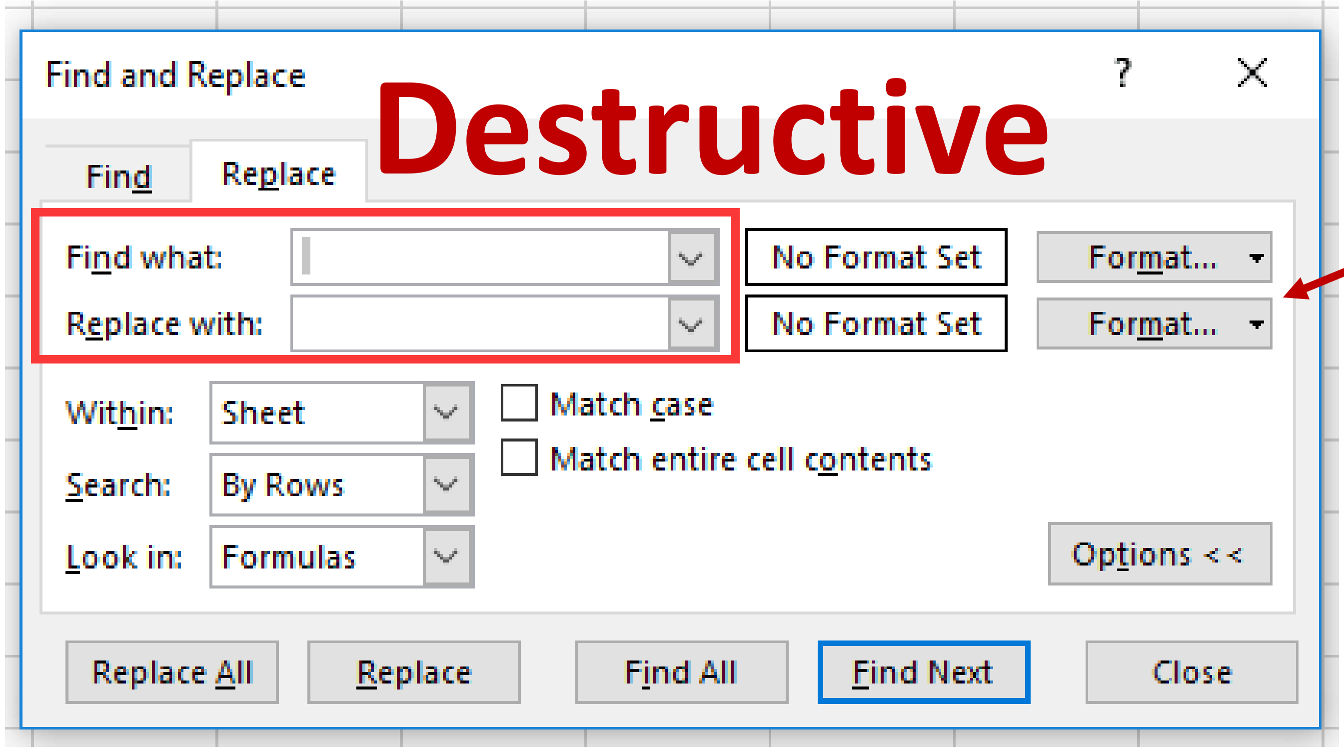
Some capitalized species.

these mean the same thing: so "spelling error"

I also have a capitalization, which R cares about ☹

White Spaces: Two Methods

- “Destructive” (editing your original data! *Eek!!!!* Protect your Raw Data!)
- “Non-Destructive/Duplicative” (Make copy of your data)



Destructive

Find and Replace

Find what:

Replace with:

Within: Sheet

Search: By Rows

Look in: Formulas

☐ Match case

☐ Match entire cell contents

If you do this on a copy of your worksheet, then it is considered non-destructive.

(It can also delete spaces you might want to keep)

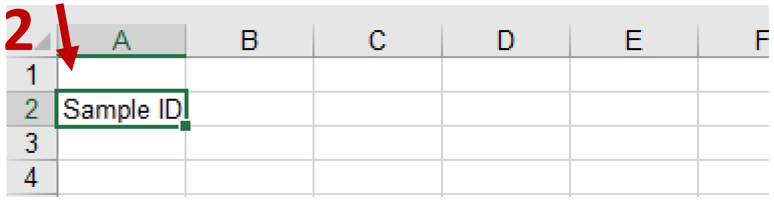
White Spaces: Non Destructive (complicated)

(but it *only* removes white spaces at the end)

1. Create New Sheet

2. =Trim(Name of original sheet!A2)
(=Trim (RAW!A2) (this will sometimes open a pop up, asking you to select the source file)

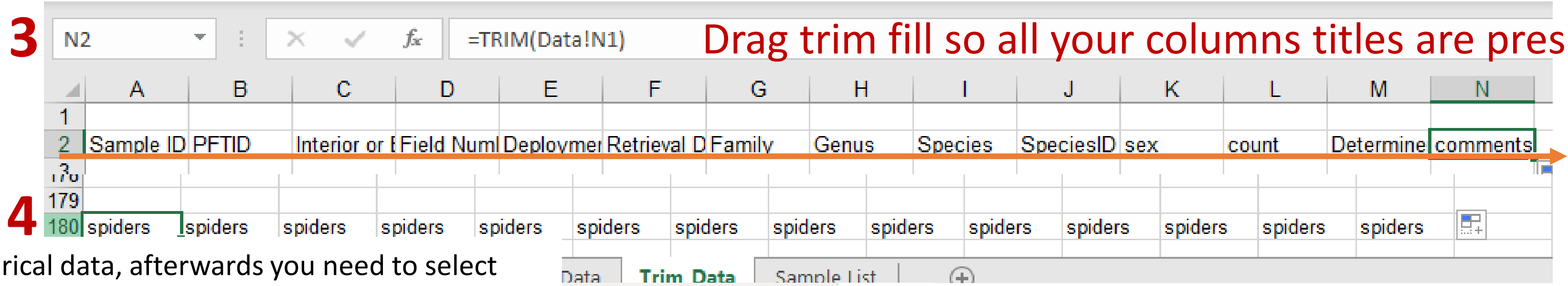
Put trim wherever your original data starts



A
Sample ID

★ This is just for ease of navigation, you don't need to do this if you remember your row/col IDs

3. In the Original sheet, hit CTRL + End, note last data cell number/letter on both cols and rows. Drag titles across accordingly.



=TRIM(Data!N1)

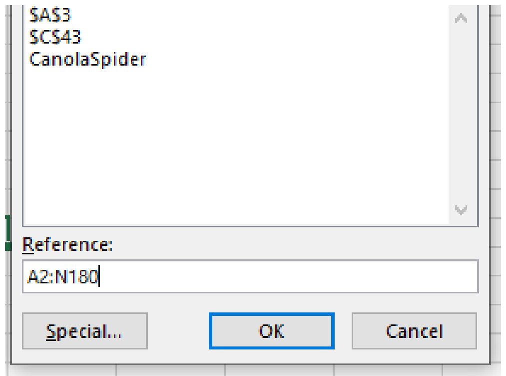
Drag trim fill so all your columns titles are present

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2	Sample ID	PFTID	Interior or E	Field Numl	Deploymer	Retrieval D	Family	Genus	Species	SpeciesID	sex	count	Determine	comments
179														
180	spiders	spiders	spiders	spiders	spiders	spiders	spiders	spiders	spiders	spiders	spiders	spiders	spiders	spiders

★ 4. In Trim sheet, go to last data row cell number and enter anything

5. CTRL + G (on trim sheet) -> Ent reference range that matches you Original Sheets Data Range. Click OK.

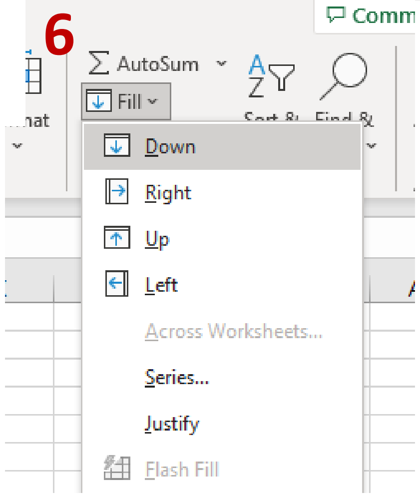
For numerical data, afterwards you need to select each column, Data-> "text to columns" and click finish
(corrects data back to numerical)



$A$3
$C$43
CanolaSpider

Reference: A2:N180

6. Under the Home tab -> Click Fill Down

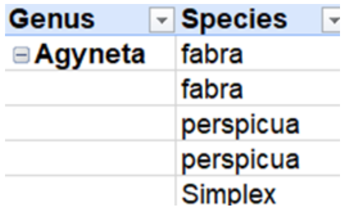


7. Copy Data -> Paste Special -> Paste values

You should now have a second datasheet without any pesky white spaces!

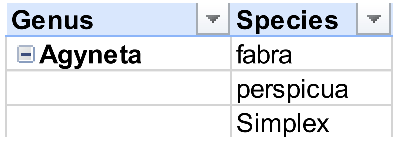
When you paste special -> 'values' it takes away the trim code

Before



Genus	Species
⊟ Agyneta	fabra
	fabra
	perspicua
	perspicua
	Simplex

After



Genus	Species
⊟ Agyneta	fabra
	perspicua
	Simplex

White Spaces: Non Destructive (complicated) (extra steps)

For numerical data, afterwards you need to select each column, Data-> “text to columns” and click finish (corrects data back to numerical)

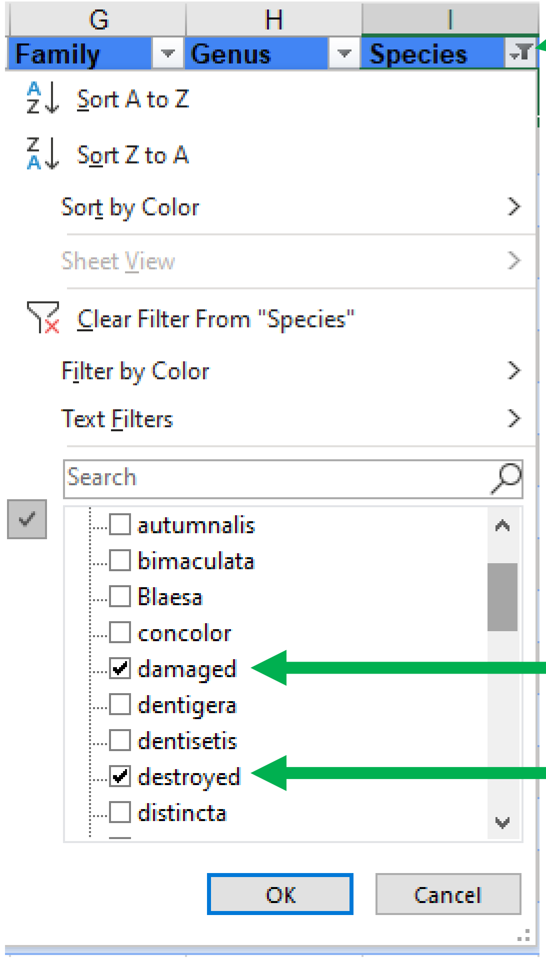
For dates that are fully numerical, it is easier to recopy the source column, trim recodes dates weird.

Typos: Use your White Space Free Table

Important:

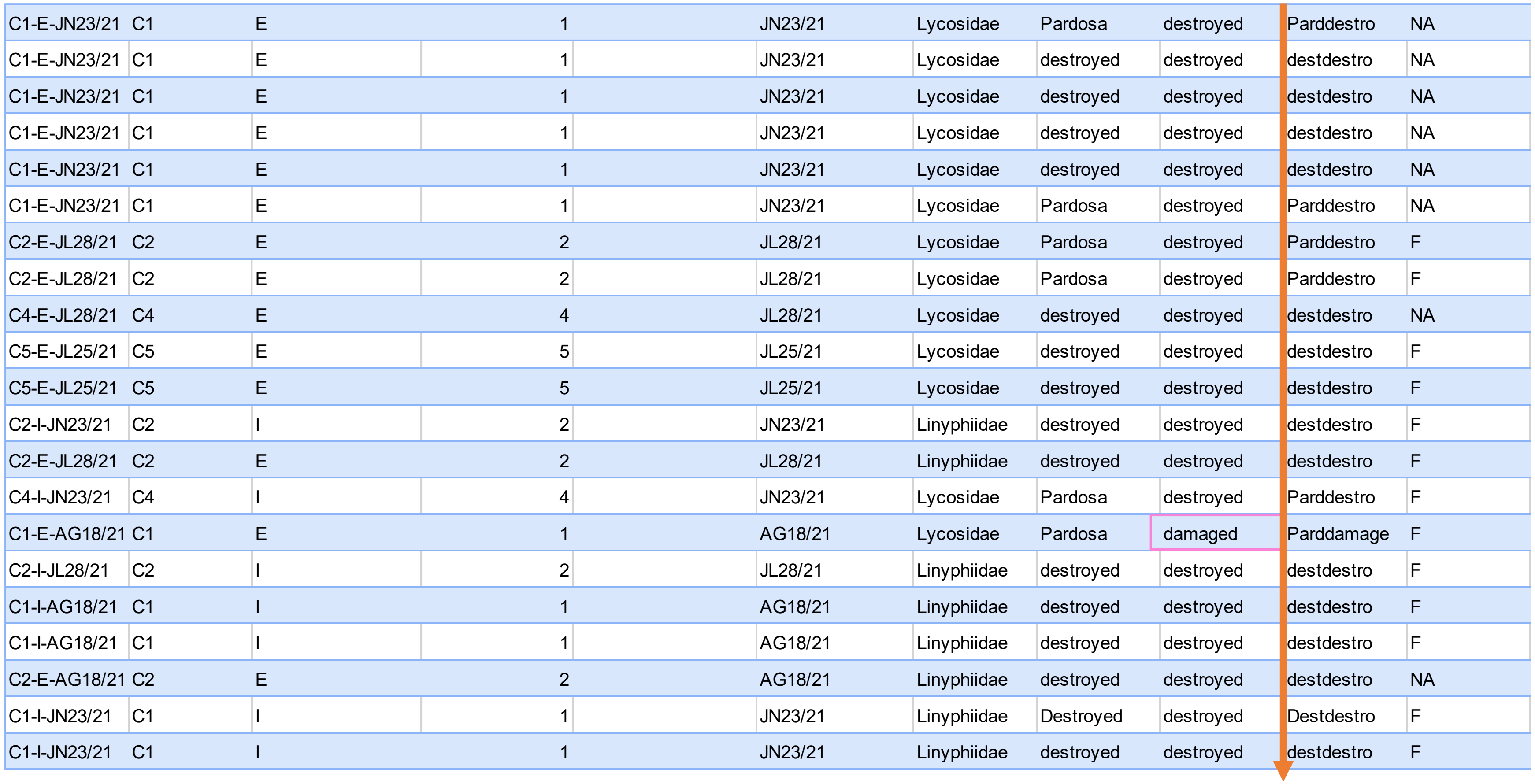
Make sure you've copied your R-Friendly Code into this column, as it was lost in the trim step

Arrows of Sorting



- ☐ autumnalis
- ☐ bimaculata
- ☐ Blaesa
- ☐ concolor
- ☑ damaged
- ☐ dentigera
- ☐ dentisetis
- ☑ destroyed
- ☐ distincta

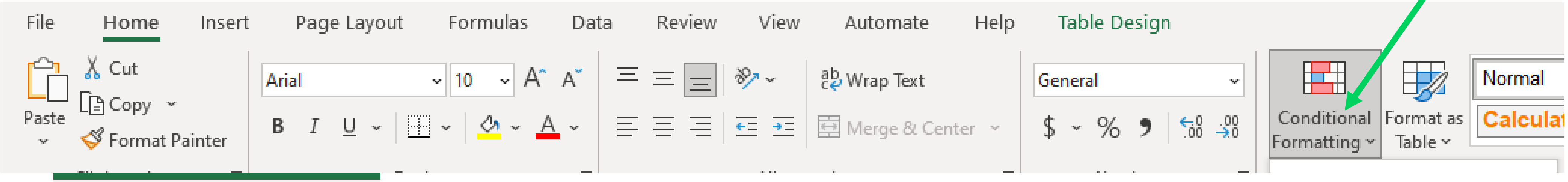
Click and drag to auto-fill



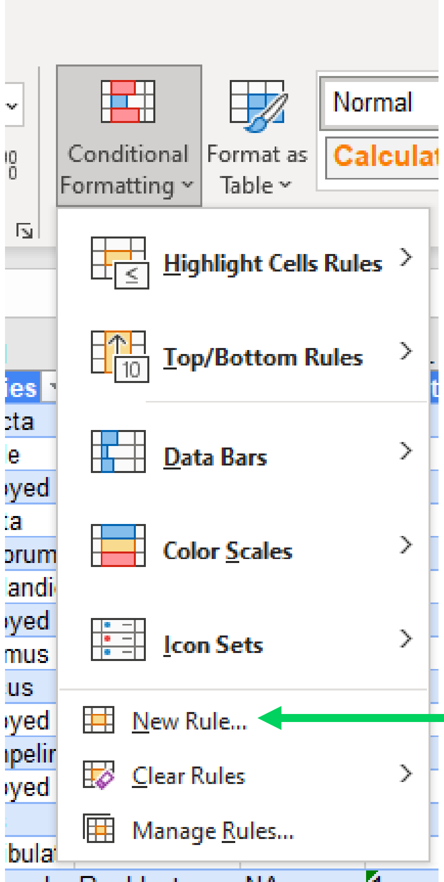
C1-E-JN23/21	C1	E	1	JN23/21	Lycosidae	Pardosa	destroyed	Parddestro	NA
C1-E-JN23/21	C1	E	1	JN23/21	Lycosidae	destroyed	destroyed	destdestro	NA
C1-E-JN23/21	C1	E	1	JN23/21	Lycosidae	destroyed	destroyed	destdestro	NA
C1-E-JN23/21	C1	E	1	JN23/21	Lycosidae	destroyed	destroyed	destdestro	NA
C1-E-JN23/21	C1	E	1	JN23/21	Lycosidae	destroyed	destroyed	destdestro	NA
C1-E-JN23/21	C1	E	1	JN23/21	Lycosidae	Pardosa	destroyed	Parddestro	NA
C2-E-JL28/21	C2	E	2	JL28/21	Lycosidae	Pardosa	destroyed	Parddestro	F
C2-E-JL28/21	C2	E	2	JL28/21	Lycosidae	Pardosa	destroyed	Parddestro	F
C4-E-JL28/21	C4	E	4	JL28/21	Lycosidae	destroyed	destroyed	destdestro	NA
C5-E-JL25/21	C5	E	5	JL25/21	Lycosidae	destroyed	destroyed	destdestro	F
C5-E-JL25/21	C5	E	5	JL25/21	Lycosidae	destroyed	destroyed	destdestro	F
C2-I-JN23/21	C2	I	2	JN23/21	Linyphiidae	destroyed	destroyed	destdestro	F
C2-E-JL28/21	C2	E	2	JL28/21	Linyphiidae	destroyed	destroyed	destdestro	F
C4-I-JN23/21	C4	I	4	JN23/21	Lycosidae	Pardosa	destroyed	Parddestro	F
C1-E-AG18/21	C1	E	1	AG18/21	Lycosidae	Pardosa	damaged	Parddamage	F
C2-I-JL28/21	C2	I	2	JL28/21	Linyphiidae	destroyed	destroyed	destdestro	F
C1-I-AG18/21	C1	I	1	AG18/21	Linyphiidae	destroyed	destroyed	destdestro	F
C1-I-AG18/21	C1	I	1	AG18/21	Linyphiidae	destroyed	destroyed	destdestro	F
C2-E-AG18/21	C2	E	2	AG18/21	Linyphiidae	destroyed	destroyed	destdestro	NA
C1-I-JN23/21	C1	I	1	JN23/21	Linyphiidae	Destroyed	destroyed	Destdestro	F
C1-I-JN23/21	C1	I	1	JN23/21	Linyphiidae	destroyed	destroyed	destdestro	F

Because you created a second data sheet at the white space stage, you have your raw data untouched and easily returned to if something goes wrong at a later step.

Additional Stuff Discussed in the meeting: Conditional Formatting



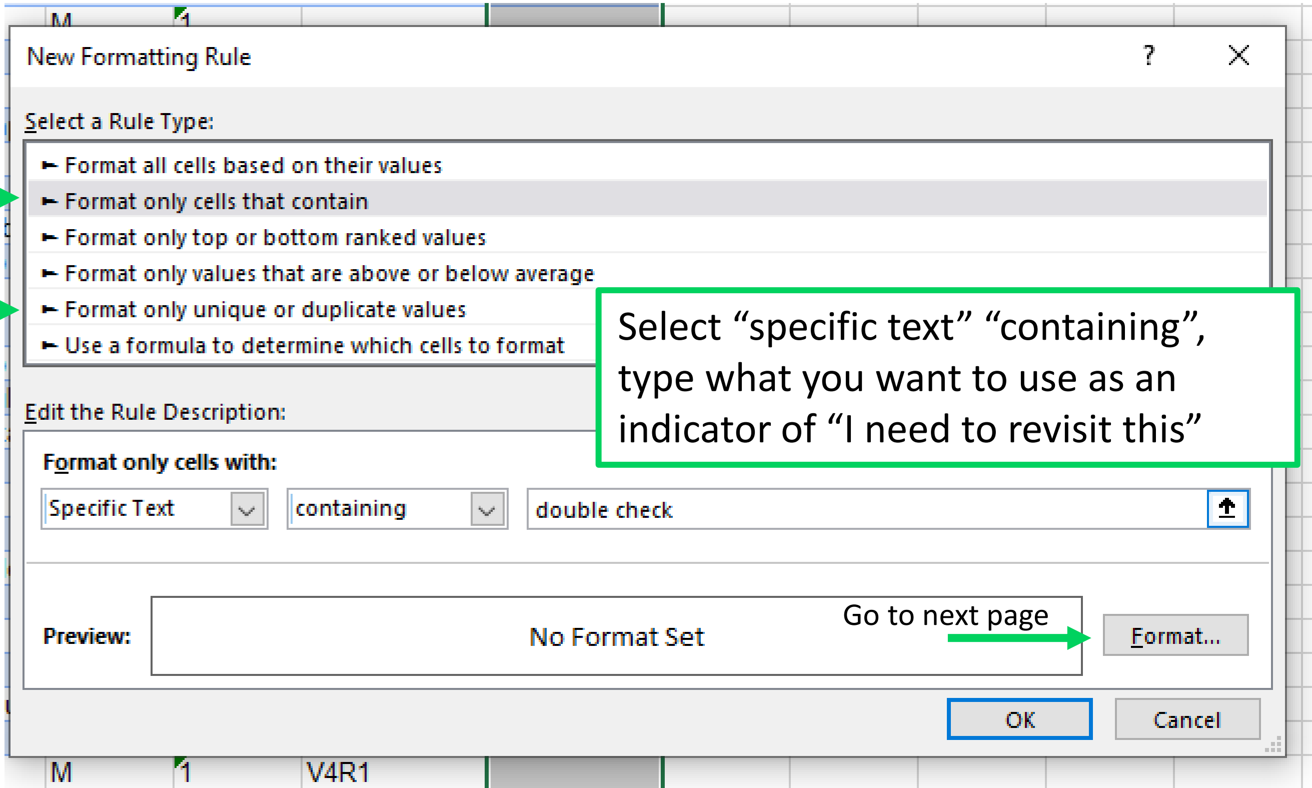
Select conditional Formatting



Select new rule

For filtering out samples that need to be revisited later using table

For filtering out duplicate sample IDs

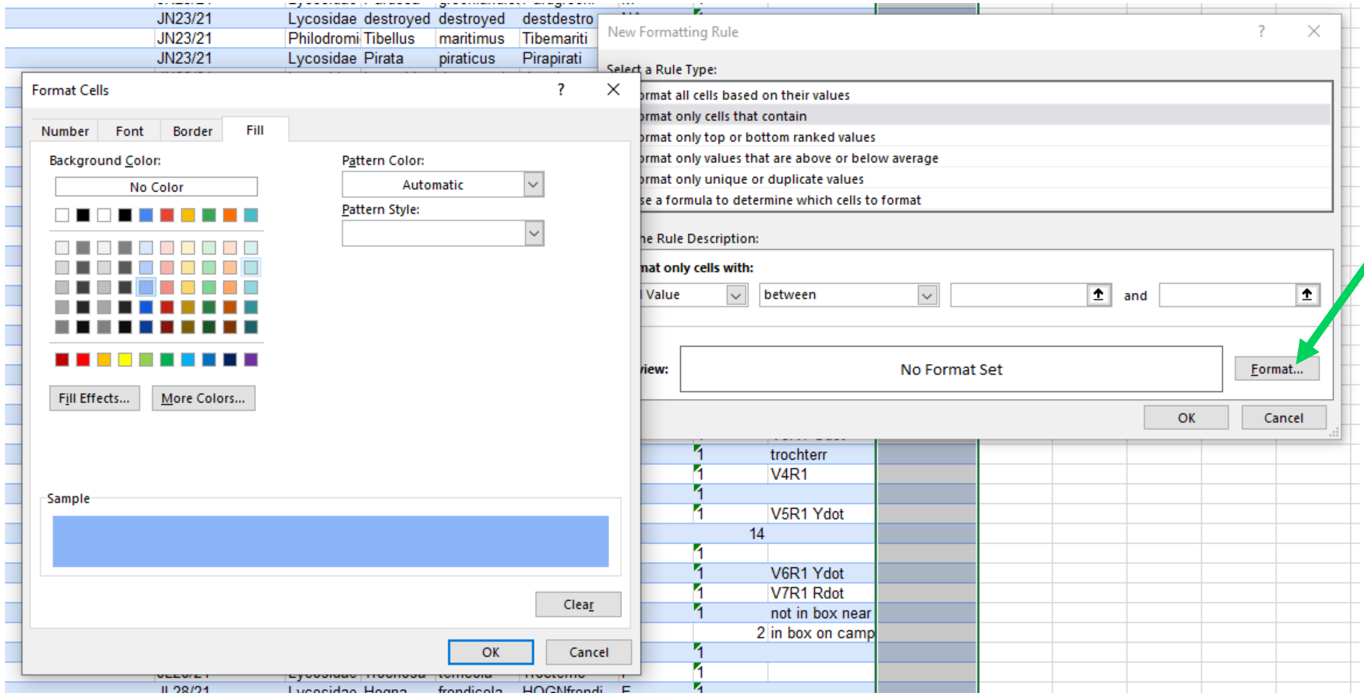


Format only cells with: Specific Text | containing | double check

Select “specific text” “containing”, type what you want to use as an indicator of “I need to revisit this”

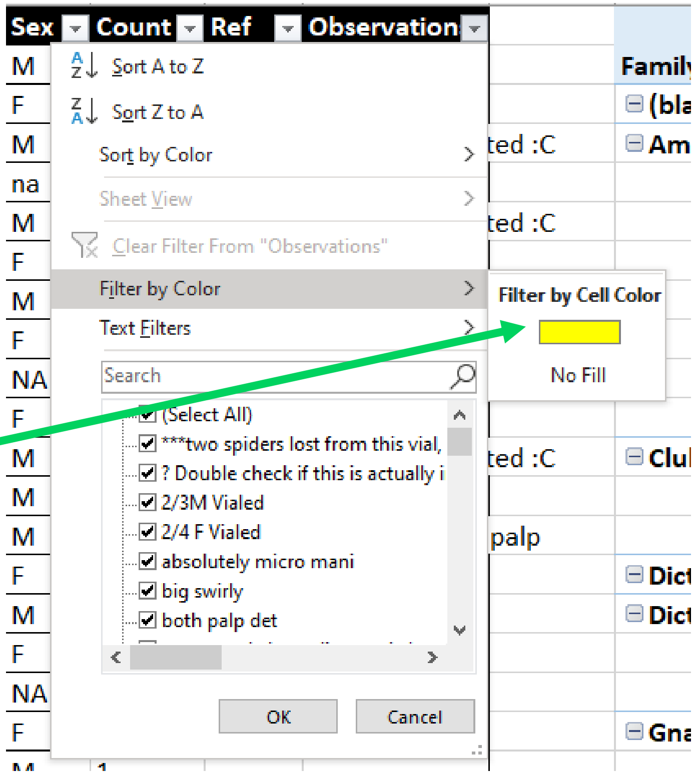
Go to next page

Conditional Formatting



Click on the formatting button, and change the fill colour of your formatting cell to something not found anywhere else in your data table.

Then, using your handy table:

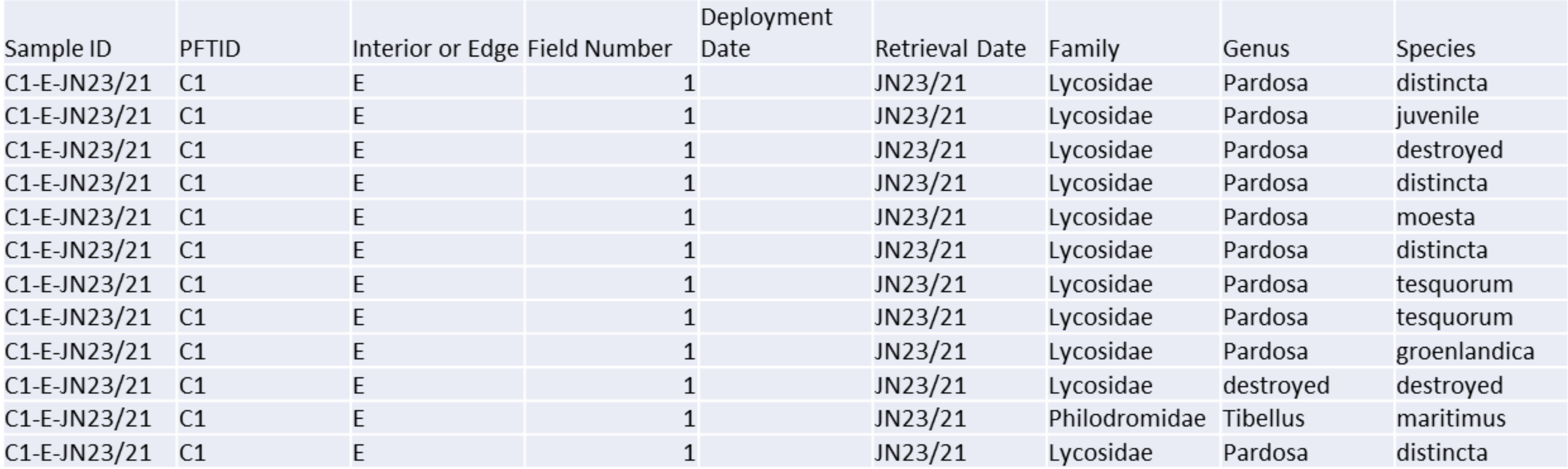


Filter by colour and ONLY things you need to revisit will be listed!

E	F	G	H	I	J	K	L
Genus	Species	sp	SpeciesID	Sex	Count	Ref	Observation
Walckenaeria	pullata	Walckenaeria pullata	Walcpullat	F	2	9-6	formerly LinySp3

In my case, I corrected an UnID'd species, but I left the column highlighted for posterity.

Duplicate Checker Using Pivot Table



Sample ID	PFTID	Interior or Edge	Field Number	Deployment Date	Retrieval Date	Family	Genus	Species
C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	distincta
C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	juvenile
C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	destroyed
C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	distincta
C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	moesta
C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	distincta
C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	tesquorum
C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	tesquorum
C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	groenlandica
C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	destroyed	destroyed
C1-E-JN23/21	C1	E	1		JN23/21	Philodromidae	Tibellus	maritimus
C1-E-JN23/21	C1	E	1		JN23/21	Lycosidae	Pardosa	distincta

This data would immediately fail the duplicate checker. I won't know if I double-entered a sample ID. Yikes!

Fixing this kind of mistake is time intensive.

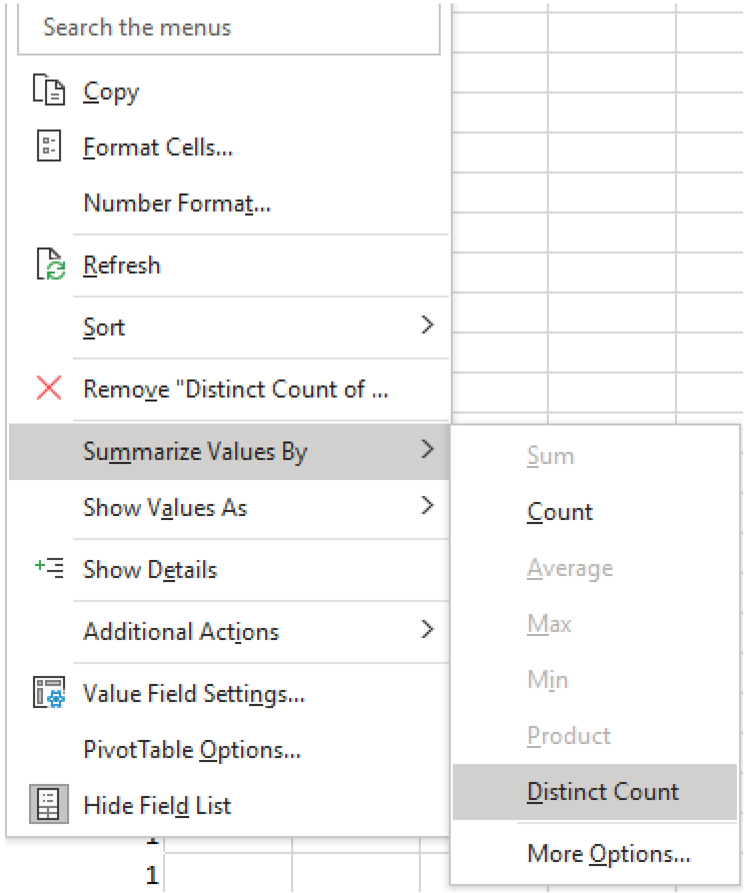
Steps one should follow for fixing this:

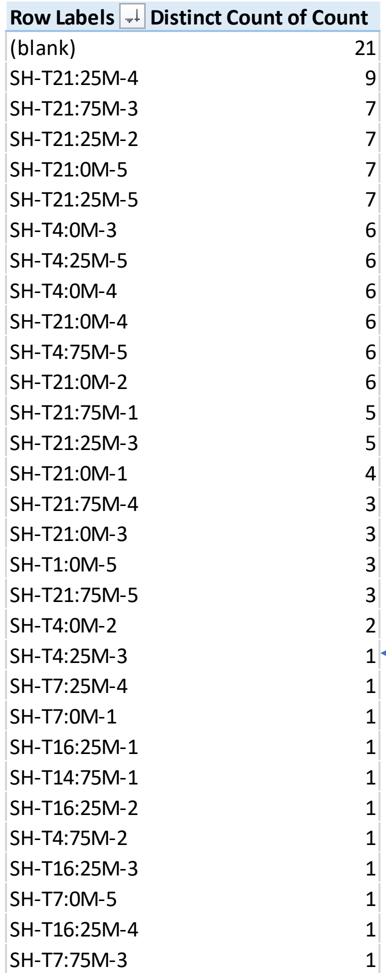
1. Sort data using helpful table arrows
2. Combine count data into single species per line/sample
3. Delete duplicated sampleID codes AFTER, NOT before.

SH-T16:0M-1.	SH-T16:0M-1	Linyphiidae	Oreonetides	vaginatus	Oreonetides vaginatus	Oreovagina	F	1
.		Thomisidae	Thomisdae	juvenile	Thomisdae juvenile	Thomjuveni	NA	2
.		Thomisidae	Xysticus	emertoni	Xysticus emertoni	Xystemerto	M	1
.		Linyphiidae	Walckenaeria	trichornis	Walckenaeria trichornis	Walctricho	F	2
.		Linyphiidae	Walckenaeria	karpinskii	Walckenaeria karpinskii	Walckarpin	M	1
.		Linyphiidae	Diplocentria	rectangulata	Diplocentria rectangulata	Diplrectan	M	1
.		Linyphiidae	Diplocentria	bidentata	Diplocentria bidentata	Diplbident	M	3
.		Linyphiidae	Diplocentria	bidentata	Diplocentria bidentata	Diplbident	F	2
.		Linyphiidae	Sicastes	truncatus	Sicastes truncatus	Sicatrunca	M	2
.		Linyphiidae	Oryphantes	aliquantulus	Oryphantes aliquantulus	Orypaliqua	M	1
.		Linyphiidae	Linyphiidae	juvenile	Linyphiidae juvenile	Linyjuveni	NA	2
SH-T17:75M-	SH-T17:75M-5	Lycosidae	Trochosa	terricola	Trochosa terricola	Trocterric	F	1
.		Lycosidae	Pardosa	uintana	Pardosa uintana	Parduintan	M	4
.		Lycosidae	Pardosa	destroyed	Pardosa destroyed	Parddestro	M	1
.		Linyphiidae	Hilaira	herniosa	Hilaira herniosa	Hilahernio	F	1
.		Linyphiidae	Walckenaeria	karpinskii	Walckenaeria karpinskii	Walckarpin	F	2

Until you are finished processing samples, it's generally better to enter data like this. The remaining trap titles can be filled in using CTRL + G *very* quickly.

Duplicate Checker





Row Labels	Distinct Count of Count
(blank)	21
SH-T21:25M-4	9
SH-T21:75M-3	7
SH-T21:25M-2	7
SH-T21:0M-5	7
SH-T21:25M-5	7
SH-T4:0M-3	6
SH-T4:25M-5	6
SH-T4:0M-4	6
SH-T21:0M-4	6
SH-T4:75M-5	6
SH-T21:0M-2	6
SH-T21:75M-1	5
SH-T21:25M-3	5
SH-T21:0M-1	4
SH-T21:75M-4	3
SH-T21:0M-3	3
SH-T1:0M-5	3
SH-T21:75M-5	3
SH-T4:0M-2	2
SH-T4:25M-3	1
SH-T7:25M-4	1
SH-T7:0M-1	1
SH-T16:25M-1	1
SH-T14:75M-1	1
SH-T16:25M-2	1
SH-T4:75M-2	1
SH-T16:25M-3	1
SH-T7:0M-5	1
SH-T16:25M-4	1
SH-T7:75M-3	1

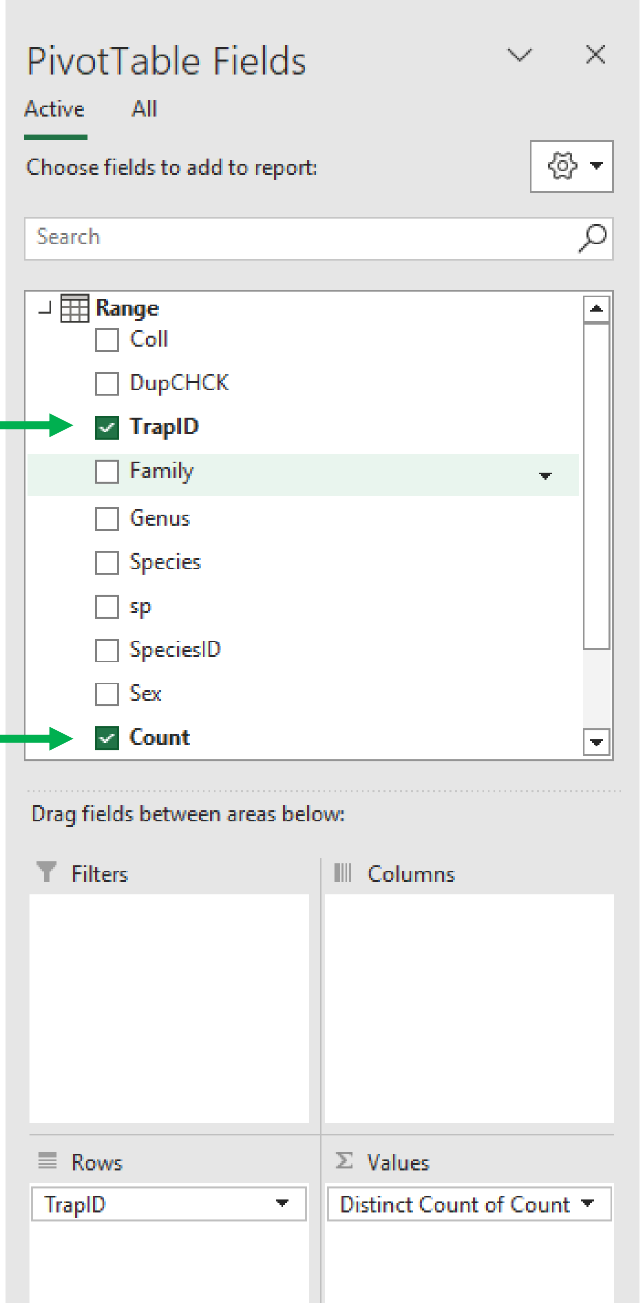
There are no actual duplicates, I went back to remove the extra sample labels.

Sample where I remembered I should use a duplicate checker.

How duplicate checker should look until the site is completed

Replicate ID/Sample ID/Test ID

Numerical Data or yes/no data (must be 0/1)



Range
☐ Coll
☐ DupCHCK
☑ **TrapID**
☐ Family
☐ Genus
☐ Species
☐ sp
☐ SpeciesID
☐ Sex
☑ **Count**

Rows: TrapID
Values: Distinct Count of Count

If you have multiple collections from the same site....

Multiple Collection Duplicate Checker

Coll	DupCHCK	TrapID	Family	Genus	Species	sp	SpeciesID	Sex	Count	Ref	Observation

This could be a
sampling date★

Eg C3E
(canola 3 edge)

=[@TrapID]&"."&[@Coll]

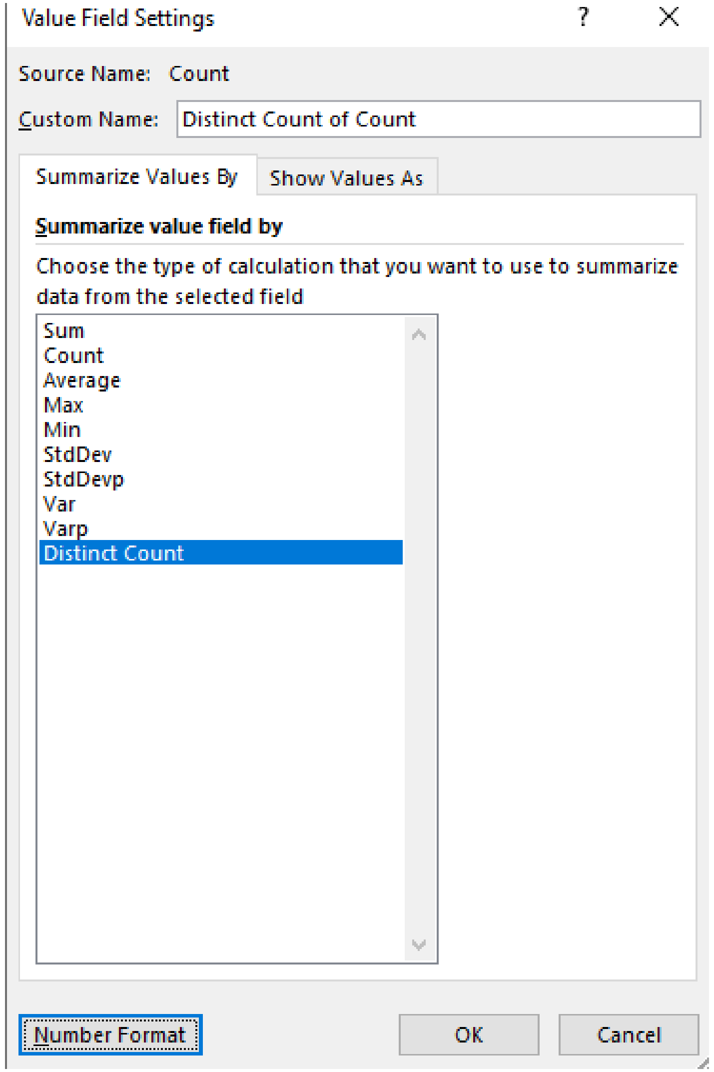
C3E.JN_20_23

On your pivot table duplicate checker, choose the DupCHCK data source instead of Trap ID.

★ In R its easier to $ call/read on plots “col1, col2, col3” so consider assigning data collection dates a more general title

The next slides are examples for Pivot table set ups very basic diversity statistical tests

Using the “summarize values” and “show values” functions on pivot tables allows you to set your data up for a lot of different statistical analyses.



Value Field Settings

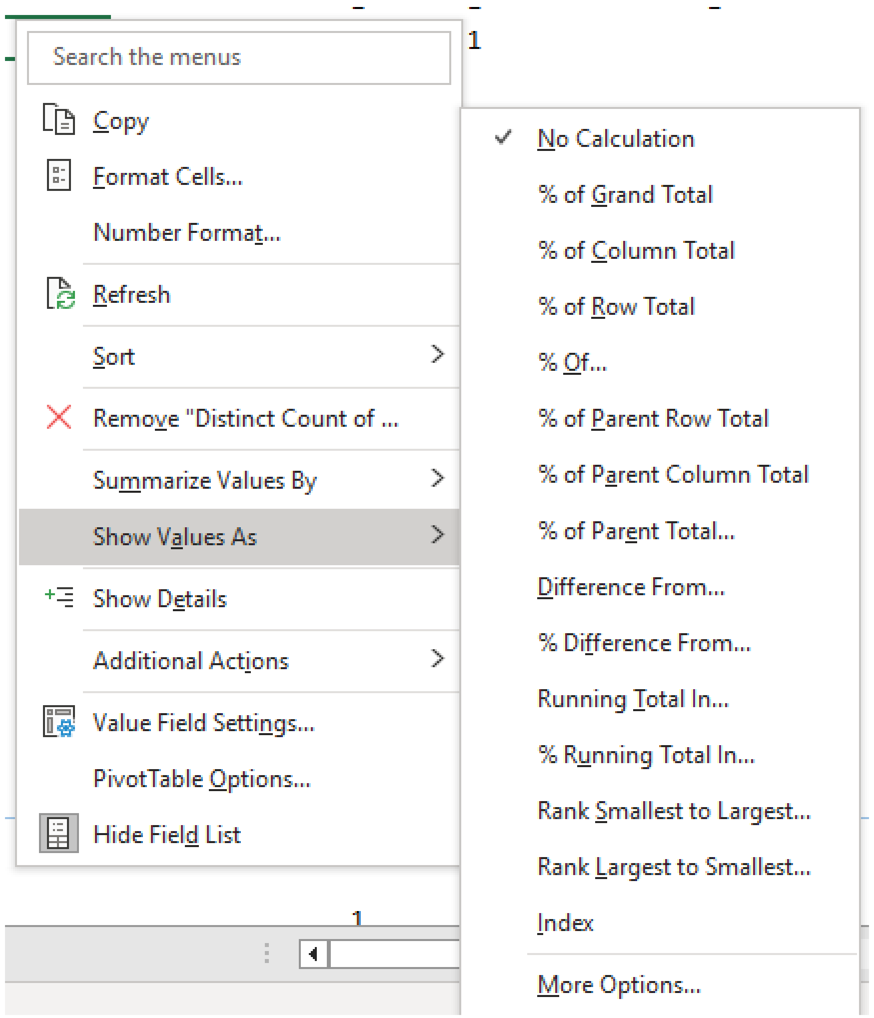
Source Name: Count

Custom Name: Distinct Count of Count

Summarize value field by

Choose the type of calculation that you want to use to summarize data from the selected field

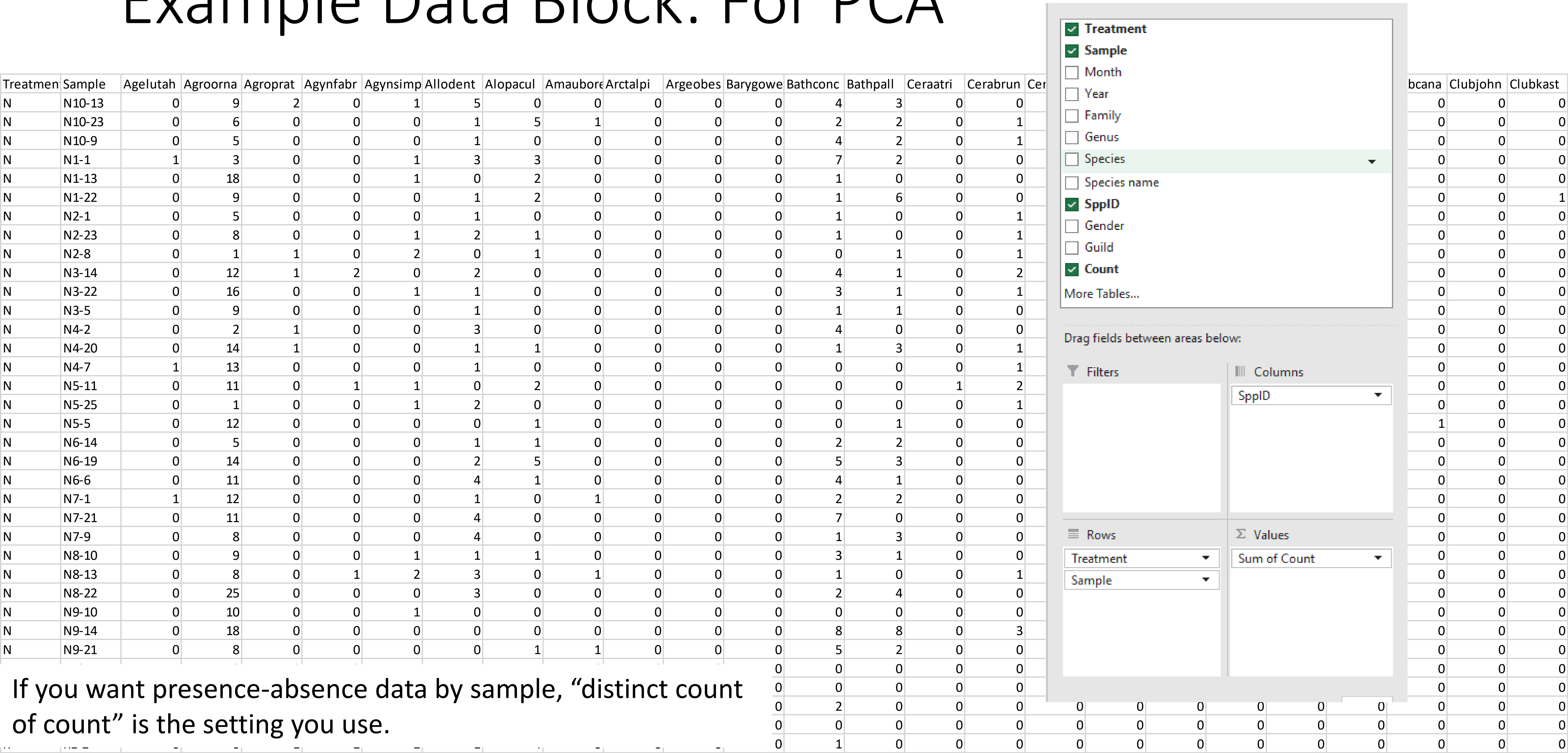
Distinct Count



You know what your data needs to do better than me!

Example Data Block: For PCA

Data Setup: Sum of Count



Treatmen	Sample	Agelutah	Agroorna	Agroprat	Agynfabr	Agynsimp	Allodent	Alopacul	Amaubore	Arctalpi	Argeobes	Barygowe	Bathconc	Bathpall	Ceraatri	Cerabrun	Cer						bcana	Clubjohn	Clubkast
N	N10-13	0	9	2	0	1	5	0	0	0	0	0	4	3	0	0							0	0	0
N	N10-23	0	6	0	0	0	1	5	1	0	0	0	2	2	0	1							0	0	0
N	N10-9	0	5	0	0	0	1	0	0	0	0	0	4	2	0	1							0	0	0
N	N1-1	1	3	0	0	1	3	3	0	0	0	0	7	2	0	0							0	0	0
N	N1-13	0	18	0	0	1	0	2	0	0	0	0	1	0	0	0							0	0	0
N	N1-22	0	9	0	0	0	1	2	0	0	0	0	1	6	0	0							0	0	1
N	N2-1	0	5	0	0	0	1	0	0	0	0	0	1	0	0	1							0	0	0
N	N2-23	0	8	0	0	1	2	1	0	0	0	0	1	0	0	1							0	0	0
N	N2-8	0	1	1	0	2	0	1	0	0	0	0	0	1	0	1							0	0	0
N	N3-14	0	12	1	2	0	2	0	0	0	0	0	4	1	0	2							0	0	0
N	N3-22	0	16	0	0	1	1	0	0	0	0	0	3	1	0	1							0	0	0
N	N3-5	0	9	0	0	0	1	0	0	0	0	0	1	1	0	0							0	0	0
N	N4-2	0	2	1	0	0	3	0	0	0	0	0	4	0	0	0							0	0	0
N	N4-20	0	14	1	0	0	1	1	0	0	0	0	1	3	0	1							0	0	0
N	N4-7	1	13	0	0	0	1	0	0	0	0	0	0	0	0	1							0	0	0
N	N5-11	0	11	0	1	1	0	2	0	0	0	0	0	0	1	2							0	0	0
N	N5-25	0	1	0	0	1	2	0	0	0	0	0	0	0	0	1							0	0	0
N	N5-5	0	12	0	0	0	0	1	0	0	0	0	0	1	0	0							1	0	0
N	N6-14	0	5	0	0	0	1	1	0	0	0	0	2	2	0	0							0	0	0
N	N6-19	0	14	0	0	0	2	5	0	0	0	0	5	3	0	0							0	0	0
N	N6-6	0	11	0	0	0	4	1	0	0	0	0	4	1	0	0							0	0	0
N	N7-1	1	12	0	0	0	1	0	1	0	0	0	2	2	0	0							0	0	0
N	N7-21	0	11	0	0	0	4	0	0	0	0	0	7	0	0	0							0	0	0
N	N7-9	0	8	0	0	0	4	0	0	0	0	0	1	3	0	0							0	0	0
N	N8-10	0	9	0	0	1	1	1	0	0	0	0	3	1	0	0							0	0	0
N	N8-13	0	8	0	1	2	3	0	1	0	0	0	1	0	0	1							0	0	0
N	N8-22	0	25	0	0	0	3	0	0	0	0	0	2	4	0	0							0	0	0
N	N9-10	0	10	0	0	1	0	0	0	0	0	0	0	0	0	0							0	0	0
N	N9-14	0	18	0	0	0	0	0	0	0	0	0	8	8	0	3							0	0	0
N	N9-21	0	8	0	0	0	0	1	1	0	0	0	5	2	0	0							0	0	0
												0	0	0	0	0							0	0	0
												0	0	0	0	0							0	0	0
												0	2	0	0	0	0	0	0	0	0	0	0	0	0
												0	0	0	0	0	0	0	0	0	0	0	0	0	0
												0	1	0	0	0	0	0	0	0	0	0	0	0	0

☑ **Treatment**
☑ **Sample**
☐ Month
☐ Year
☐ Family
☐ Genus
☐ Species
☐ Species name
☑ **SppID**
☐ Gender
☐ Guild
☑ **Count**

Columns: SppID

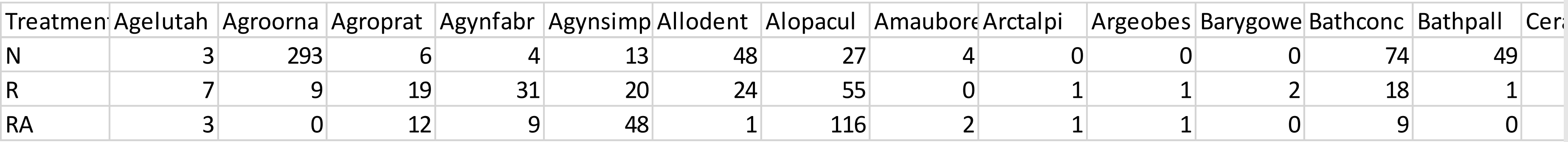
Rows: Treatment, Sample

Values: Sum of Count

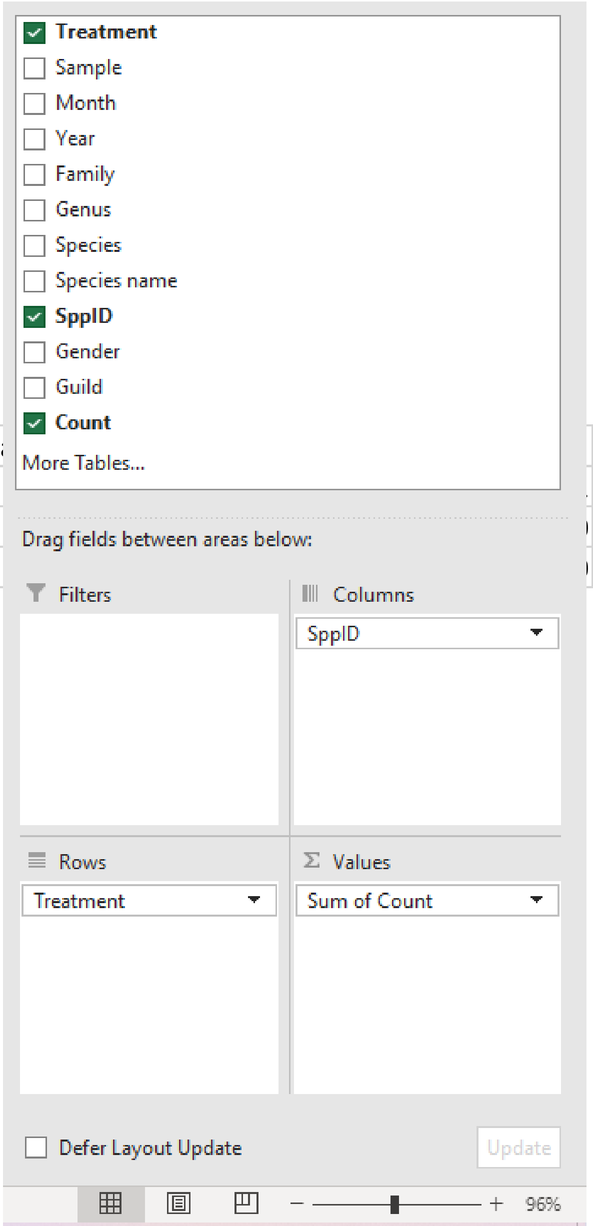
If you want presence-absence data by sample, “distinct count of count” is the setting you use.

Example Data Block: Rarefaction

Data Setup: Sum of Count



Treatmen	Agelutah	Agroorna	Agroprat	Agynfabr	Agynsimp	Allodent	Alopacul	Amaubore	Arctalpi	Argeobes	Barygowe	Bathconc	Bathpall	Cer	Cerasimi
N	3	293	6	4	13	48	27	4	0	0	0	74	49		0
R	7	9	19	31	20	24	55	0	1	1	2	18	1		0
RA	3	0	12	9	48	1	116	2	1	1	0	9	0		2



☑ **Treatment**
☐ Sample
☐ Month
☐ Year
☐ Family
☐ Genus
☐ Species
☐ Species name
☑ **SppID**
☐ Gender
☐ Guild
☑ **Count**
More Tables...

Drag fields between areas below:

Filters

Columns: SppID

Rows: Treatment

Values: Sum of Count

☐ Defer Layout Update

(this chart goes on until column DO so I couldn't include the whole thing)

Importing into R

- library(readxl) # to read excel files

File Name

Worksheet Name

```
# Read data, read version Dec112020
spider.dat <- as.data.frame(read_excel('SpidDat2020.xlsx', 'RareDat',
range='A1:DN4’))
# Tells R how to read the row names in your data frame, otherwise R will
treat your vertical row names as data and throw errors.
rownames(spider.dat) <- spider.dat[,1]; spider.dat <- spider.dat[,-1]
```

Cells to import from sheet. If you get this wrong R *will* yell at you.

- Calling ‘SpidDat2020.xlsx’ will only work if your working directory is set up correctly. Otherwise you need to enter the full directory: ‘C:/Users/name_azlkirs/Documents/499/ESP2.xlsx’