

Wrangling and Tidying for the infrequent useR

Kim Cressman

Grand Bay NERR

kimberly.cressman@msstate.edu

May 19-20, 2022

Realistic expectations

What I hope you get out of this workshop

- Go write all your code from scratch, from memory
- Know **what's possible**
- **Identify** what you need to do
- Know enough **vocabulary** and help-finding techniques to **look up** what you need

Intro to Datasets

What kind of data are we working with?

- Fish monitoring
- Vegetation monitoring
- Water quality data
- *C. elegans* toxicity data

More detail here:

https://github.com/swmpkim/2022_infrequentUseR_wrangling/tree/main/data

What will we be doing with it?

- Note: We will be going over all of this in much more detail over the next 1.5 days! This is just a preview!

Vegetation Data

3 files: one year of data in each

	A	B	C	D	E	F	G	H	I	J	
1	Reserve	Date	SiteID	TransectID	PlotID	Lat	Long	Orthometr	Species	Cover	
2	GND	7/26/2018	spal	1	1	30.36248	-88.4139	0.22874	Juncus roemerianus	20	
3	GND	7/26/2018	spal	1	1	30.36248	-88.4139	0.22874	Spartina alterniflora	10	
4	GND	7/26/2018	spal	1	2	30.36236	-88.4138	0.18774	Juncus roemerianus	20	
5	GND	7/26/2018	spal	1	2	30.36236	-88.4138	0.18774	Spartina alterniflora	5	
6	GND	7/26/2018	spal	1	3	30.36224	-88.4137	0.18604	Juncus roemerianus	0	
7	GND	7/26/2018	spal	1	3	30.36224	-88.4137	0.18604	Spartina alterniflora	50	
8	GND	7/26/2018	spal	1	4	30.36212	-88.4136	0.20855	Juncus roemerianus	55	
9	GND	7/26/2018	spal	1	4	30.36212	-88.4136	0.20855	Spartina alterniflora	2.5	
10	GND	7/26/2018	spal	1	5	30.36199	-88.4135	0.21419	Juncus roemerianus	30	
11	GND	7/26/2018	spal	1	5	30.36199	-88.4135	0.21419	Spartina alterniflora	0	
12	GND	7/26/2018	spal	1	6	30.36186	-88.4134	0.24428	Juncus roemerianus	40	
13	GND	7/26/2018	spal	1	6	30.36186	-88.4134	0.24428	Spartina alterniflora	0	
14	GND	7/26/2018	spal	1	7	30.36173	-88.4133	0.21105	Juncus roemerianus	25	
15	GND	7/26/2018	spal	1	7	30.36173	-88.4133	0.21105	Spartina alterniflora	0	
16	GND	7/26/2018	spal	1	8	30.36162	-88.4133	0.19436	Juncus roemerianus	30	

- File names start with 'GNDveg'
- We will mostly be working with Species and Cover

Read in and bind multiple files

2018

	A	B	C	D	E	F	G	H	I	J
1	Reserve	Date	SiteID	TransectID	PlotID	Lat	Long	Orthometr	Species	Cover
2	GND	7/26/2018	spal	1	1	30.36248	-88.4139	0.22874	Juncus roemerianus	20
3	GND	7/26/2018	spal	1	1	30.36248	-88.4139	0.22874	Spartina alterniflora	10
4	GND	7/26/2018	spal	1	2	30.36236	-88.4138	0.18774	Juncus roemerianus	20
5	GND	7/26/2018	spal	1	2	30.36236	-88.4138	0.18774	Spartina alterniflora	5
6	GND	7/26/2018	spal	1	3	30.36224	-88.4137	0.18604	Juncus roemerianus	0
7	GND	7/26/2018	spal	1	3	30.36224	-88.4137	0.18604	Spartina alterniflora	50

2019

	A	B	C	D	E	F	G	H	I	J
1	Reserve	Date	SiteID	TransectID	PlotID	Lat	Long	Orthometr	Species	Cover
2	GND	8/13/2019	juro mid	1	1	30.39773	-88.4125	0.34721	Borrighia frutescens	0
3	GND	8/13/2019	juro mid	1	1	30.39773	-88.4125	0.34721	Distichlis spicata	0
4	GND	8/13/2019	juro mid	1	1	30.39773	-88.4125	0.34721	Juncus roemerianus	5
5	GND	8/13/2019	juro mid	1	1	30.39773	-88.4125	0.34721	Spartina patens	0
6	GND	8/13/2019	juro mid	1	1	30.39773	-88.4125	0.34721	Other	25
7	GND	8/13/2019	juro mid	1	2	30.39773	-88.4128	0.32548	Borrighia frutescens	0
8	GND	8/13/2019	juro mid	1	2	30.39773	-88.4128	0.32548	Distichlis spicata	0

2020

	A	B	C	D	E	F	G	H	I	J
1	Reserve	Date	SiteID	TransectID	PlotID	Lat	Long	Orthometr	Species	Cover
2	GND	8/6/2020	juro mid	1	1	30.39773	-88.4125	0.28723	Borrighia frutescens	0
3	GND	8/6/2020	juro mid	1	1	30.39773	-88.4125	0.28723	Distichlis spicata	0
4	GND	8/6/2020	juro mid	1	1	30.39773	-88.4125	0.28723	Juncus roemerianus	15
5	GND	8/6/2020	juro mid	1	1	30.39773	-88.4125	0.28723	Spartina patens	0
6	GND	8/6/2020	juro mid	1	1	30.39773	-88.4125	0.28723	Other	25
7	GND	8/6/2020	juro mid	1	2	30.39772	-88.4128	0.25081	Borrighia frutescens	0
8	GND	8/6/2020	juro mid	1	2	30.39772	-88.4128	0.25081	Distichlis spicata	0

Key packages and functions

dplyr

- `bind_rows()`

▶ veg_all	1077 obs. of 10 variables
▶ veg2018	359 obs. of 10 variables
▶ veg2019	359 obs. of 10 variables
▶ veg2020	359 obs. of 10 variables

Key packages and functions

dplyr
• bind_rows()

	A	B	C	D	E	F	G	H	I	J
1	Reserve	Date	SiteID	TransectID	PlotID	Lat	Long	Orthometr	Species	Cover
2	GND	7/26/2018	spal	1	1	30.36248	-88.4139	0.22874	Juncus roemerianus	20
3	GND	7/26/2018	spal	1	1	30.36248	-88.4139	0.22874	Spartina alterniflora	10
4	GND	7/26/2018	spal	1	2	30.36236	-88.4138	0.18774	Juncus roemerianus	20
5	GND	7/26/2018	spal	1	2	30.36236	-88.4138	0.18774	Spartina alterniflora	5
6	GND	7/26/2018	spal	1	3	30.36224	-88.4137	0.18604	Juncus roemerianus	0
7	GND	7/26/2018	spal	1	3	30.36224	-88.4137	0.18604	Spartina alterniflora	50
8	GND	7/26/2018	spal	1	4	30.36212	-88.4136	0.20855	Juncus roemerianus	55
1071	GND	8/12/2020	spal	3	6	30.3617	-88.4139	0.18954	Juncus roemerianus	30
1072	GND	8/12/2020	spal	3	6	30.3617	-88.4139	0.18954	Spartina alterniflora	5
1073	GND	8/12/2020	spal	3	7	30.36159	-88.4138	0.14866	Juncus roemerianus	50
1074	GND	8/12/2020	spal	3	7	30.36159	-88.4138	0.14866	Spartina alterniflora	2.5
1075	GND	8/12/2020	spal	3	8	30.36147	-88.4138	0.16786	Juncus roemerianus	50
1076	GND	8/12/2020	spal	3	8	30.36147	-88.4138	0.16786	Spartina alterniflora	2.5
1077	GND	8/12/2020	spal	3	9	30.36131	-88.4137	0.17727	Juncus roemerianus	40
1078	GND	8/12/2020	spal	3	9	30.36131	-88.4137	0.17727	Spartina alterniflora	5

Group and summarize data

- e.g. min, max, and median cover by species, by site and/or year

SiteID	Species	min_Cover	mean_Cover	median_Cover	max_Cover
juro low	Juncus roemerianus	0	25.14	25.0	60.0
juro mid	Juncus roemerianus	0	21.77	22.5	65.0
spal	Juncus roemerianus	0	18.16	20.0	55.0
spal	Spartina alterniflora	0	16.67	10.0	60.0
juro mid	Other	0	7.38	2.5	65.0
juro low	Spartina alterniflora	0	4.74	0.0	60.0
juro mid	Borrchia frutescens	0	2.83	2.5	15.0
juro mid	Distichlis spicata	0	1.40	0.0	20.0
juro mid	Spartina patens	0	0.73	0.0	35.0
juro low	Distichlis spicata	0	0.26	0.0	2.5

Key packages and functions

dplyr

- `group_by()`
- `summarize()`

We can go from 1077 rows to 15 by grouping and summarizing!

Pivot data to a wider format

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Reserve	Date	SiteID	TransectID	PlotID	Lat	Long	Orthome	Juncus roemerianus	Spartina alterniflora	orrichia frutescens	Distichlis spicata	Spartina patens	Other	
2	GND	7/26/2018	spal	1	1	30.36248	-88.4139	0.2287							
3	GND	7/26/2018	spal	1	2	30.36236	-88.4138	0.18774	20	5					
4	GND	7/26/2018	spal	1	3	30.36224	-88.4137	0.18604	0	30					
5	GND	7/26/2018	spal	1	4	30.36212	-88.4136	0.20855	55	2.5					
6	GND	7/26/2018	spal	1	5	30.36199	-88.4135	0.21419	30	0					
7	GND	7/26/2018	spal	1	6	30.36186	-88.4134	0.24428	40	0					
	A	B	C	D	E	F	G	H	I	J					
1	Reserve	Date	SiteID	TransectID	PlotID	Lat	Long	Orthometr	Species	Cover					
2	GND	7/26/2018	spal	1	1	30.36248	-88.4139	0.22874	Juncus roemerianus	20		2.5			
3	GND	7/26/2018	spal	1	1	30.36248	-88.4139	0.22874	Spartina alterniflora	10		30			
4	GND	7/26/2018	spal	1	2	30.36236	-88.4138	0.18774	Juncus roemerianus	20		20			
5	GND	7/26/2018	spal	1	2	30.36236	-88.4138	0.18774	Spartina alterniflora	5		10			
6	GND	7/26/2018	spal	1	3	30.36224	-88.4137	0.18604	Juncus roemerianus	0		20			
7	GND	7/26/2018	spal	1	3	30.36224	-88.4137	0.18604	Spartina alterniflora	50		2.5			
8	GND	7/26/2018	spal	1	4	30.36212	-88.4136	0.20855	Juncus roemerianus	55					
9	GND	7/26/2018	spal	1	4	30.36212	-88.4136	0.20855	Spartina alterniflora	2.5					
10	GND	7/26/2018	spal	1	5	30.36199	-88.4135	0.21419	Juncus roemerianus	30					
11	GND	7/26/2018	spal	1	5	30.36199	-88.4135	0.21419	Spartina alterniflora	0					
12	GND	7/26/2018	spal	1	6	30.36186	-88.4134	0.24428	Juncus roemerianus	40					
13	GND	7/26/2018	spal	1	6	30.36186	-88.4134	0.24428	Spartina alterniflora	0					
14	GND	7/26/2018	spal	1	7	30.36173	-88.4133	0.21105	Juncus roemerianus	25					
15	GND	7/26/2018	spal	1	7	30.36173	-88.4133	0.21105	Spartina alterniflora	0					
16	GND	7/26/2018	spal	1	8	30.36162	-88.4133	0.19436	Juncus roemerianus	30					

Key packages and functions

tidyr
• pivot_wider()

Water Quality data

2 files: ~ one month of data in each

	A	B	C	D	E	F	G	H	I	J	K	
1	KorEXO MEASUREMENT DATA FILE EXPORT											
2												
3	FILE CREATED: #####											
4												
5				MEAN VALUE	0.598	93.7	8.76	0.709	19.2	30.703	5.14	
6				STANDARD DE	0.237	10.9	1.23	0.343	5.41	8.392	4	
7												
8				SENSOR SERIAL	14E10290	13A10130	13A10130	14E10290	19G10055	19G10055	13G10030	14
9	Date (MM/DD)	Time (HH:mm)	Time (Frac)	Site Name	Depth m	ODO % sat	ODO mg/L	Pressure p	Sal psu	SpCond m	Turbidity F W	
10	1/24/2022	11:45:00	0	GNDBCWQ	0.542	99.5	10.05	0.629	20.2	32.493	47.2	
11	1/24/2022	12:00:00	0	GNDBCWQ	0.53	98.6	9.81	0.612	21.81	34.824	7	
12	1/24/2022	12:15:00	0	GNDBCWQ	0.548	102.8	10.08	0.638	22.81	36.262	4.83	
13	1/24/2022	12:30:00	0	GNDBCWQ	0.557	102.3	10.11	0.651	21.62	34.538	4.13	
14	1/24/2022	12:45:00	0	GNDBCWQ	0.553	100.2	10.01	0.645	20.53	32.943	3.47	
15	1/24/2022	13:00:00	0	GNDBCWQ	0.549	100.7	10.04	0.639	20.54	32.962	3.26	
16	1/24/2022	13:15:00	0	GNDBCWQ	0.569	101.8	10.13	0.668	20.9	33.485	2.96	
17	1/24/2022	13:30:00	0	GNDBCWQ	0.582	103	10.17	0.687	21.19	33.895	3.14	
18	1/24/2022	13:45:00	0	GNDBCWQ	0.578	104.1	10.21	0.68	21.32	34.077	2.82	
19	1/24/2022	14:00:00	0	GNDBCWQ	0.579	104.6	10.25	0.682	21.54	34.388	2.98	
20	1/24/2022	14:15:00	0	GNDBCWQ	0.578	104.8	10.26	0.68	21.35	34.102	3.08	
21	1/24/2022	14:30:00	0	GNDBCWQ	0.578	105.2	10.29	0.68	21.34	34.096	2.99	
22	1/24/2022	14:45:00	0	GNDBCWQ	0.575	104.6	10.25	0.676	20.74	33.215	3.19	
23	1/24/2022	15:00:00	0	GNDBCWQ	0.577	105	10.28	0.679	20.87	33.405	3.1	
24	1/24/2022	15:15:00	0	GNDBCWQ	0.63	105.7	10.33	0.755	21.13	33.776	2.98	
25	1/24/2022	15:30:00	0	GNDBCWQ	0.655	108.2	10.43	0.793	22.53	35.804	3.51	
26	1/24/2022	15:45:00	0	GNDBCWQ	0.645	108.3	10.44	0.778	22.29	35.458	3.58	
27	1/24/2022	16:00:00	0	GNDBCWQ	0.643	108.1	10.4	0.775	22.55	35.832	3.81	

- File names start with 'GNDBCWQ'
- Exported from KOR software

Read in a file with lots of rows at the top

- When you know how many rows to skip
- When you want to **identify a row** to start on

Key packages and functions

base

- `read.csv(skip = ____)`

stringr

- `str_which()`

1	KorEXO MEASUREMENT DATA FILE EXPORT																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
2																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
3	FILE CREATED: #####																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
4																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
5	MEAN VALUE 0.598																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
6	STANDARD DE 0.237 10.9 1.23 0.343 5.41 8.392																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
7																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
8	SENSOR SERIAL 14E10290 13A10130 13A10130 14E10290 19G10055 19G10055																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
9	Date (MM/DD)	Time (HH:mm)	Time (Frac)	Site Name	Depth m	ODO % sat	ODO mg/L	Pressure p	Sal psu	SpCond m																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	</

Bind files together, group, and summarize (Bonus Content)

- e.g. create daily or monthly averages
- Note, this also involves working with dates!
- And selection of multiple columns to summarize

Key packages
and functions

dplyr

- `bind_rows()`
- `group_by()`
- `summarize()`

lubridate

	Date	Depth_m_mean	Depth_m_min	Depth_m_max	DO_pct_mean	DO_pct_min	DO_pct_max	DO_mgl_mean			
1	2022-01-24	0.5909184	0.526	0.675	105.71224	98.6	110.6	10.2585			
2	2022-01-25	0.6244688	0.462	0.770	100.71771	93.2	104.1	9.4942			
3	2022-01-26	0.6558437	0.369	0.951	93.36562	82.1	105.6	9.046875	7.90	10.01	
4	2022-01-27	0.6423646	0.351	0.920	90.75521	79.4	102.6	8.840938	7.93	9.77	
5	2022-01-28	0.4515312	0.131	0.789	88.69167	78.9	99.6	8.752917	7.82	9.72	
6	2022-01-29	0.4813125	0.180	0.835	99.10729	83.0	111.1	11.165000	8.74	16.93	
7	2022-01-30	0.4861563	0.162	0.893	99.06875	87.6	107.9	9.921875	9.41	10.42	
8	2022-01-31	0.5421458	0.135	1.104	103.82083	89.1	120.9	9.857083	8.79	10.99	
9	2022-02-01	0.7070937	0.330	1.203	100.79063	81.3	116.1	9.140833	7.64	10.58	

Fish monitoring data

3 files: different types of data

	A	B	C	D	E
1	site	habitat_type	lat	long	location
2	11	erosional edge	30.37163	-88.4438	Bayou Cumbest
3	14	erosional edge	30.3557	-88.4495	Pt aux Chens Bay
4	2	seagrass	30.38508	-88.4022	Middle Bay
5	5	seagrass	30.36205	-88.3977	Grand Bay
6	6	erosional edge	30.34905	-88.3973	Grand Battures
7	8	seagrass	30.35493	-88.4106	Jose Bay

one-to-many

	A	B	C	D	E	F	G
1	collection_id	site	season	year_sampled	salinity_ppt	do_mgl	water_temp_c
2	NFM08-142	2	Winter	2008	18.4	8.24	14
3	NFM08-143	5	Winter	2008	17.3	7.98	14.5
4	NFM08-146	6	Winter	2008	17.8	8.68	13.9
5	NFM08-148	8	Winter	2008	19.3	8.52	15.2
6	NFM08-151	11	Winter	2008	18.1	7.27	17.1
7	NFM08-154	14	Winter	2008	19.6	9.12	18.3
8	NFM08-156	2	Spring	2008	17.4	6.15	27.4
9	NFM08-157	5	Spring	2008	18.7	5.8	28.2
10	NFM08-160	6	Spring	2008	18.2	6.17	30
11	NFM08-162	8	Spring	2008	18.7	7.16	29.1
12	NFM08-165	11	Spring	2008	12.9	5.92	31.9
13	NFM08-168	14	Spring	2008	16.7	7.72	31.9
14	NFM08-169	2	Summer	2008	18.8	3.42	29.6
15	NFM08-170	5	Summer	2008	19.2	4.12	28.6
16	NFM08-173	6	Summer	2008	19.2	5.78	29.8
17	NFM08-175	8	Summer	2008	20.8	5.07	30.2
18	NFM08-178	11	Summer	2008	10.4	4.29	31.9
19	NFM08-180	14	Summer	2008	20.5	5.81	31.5

File names start with 'GNDfish'

- _sites
- _wq
- _lengths

3 files: different types of data

	A	B	C	D	E	F	G
1	collection_id	site	season	year_sampled	salinity_ppt	do_mgl	water_temp_c
2	NFM08-142	2	Winter	2008	18.4	8.24	14
3	NFM08-143	3	Winter	2008	17.3	7.98	14.5
4	NFM08-146	6	Winter	2008	17.8	8.68	13.9
5	NFM08-148	8	Winter	2008	19.3	8.52	15.2
6	NFM08-151	11	Winter	2008	18.1	7.27	17.1
7	NFM08-154	14	Winter	2008	19.6	9.12	18.3
8	NFM08-156	2	Spring	2008	17.4	6.15	27.4
9	NFM08-157	3	Spring	2008	18.7	5.8	28.2
10	NFM08-160	6	Spring	2008	18.2	6.17	30
11	NFM08-162	8	Spring	2008	18.7	7.16	29.1
12	NFM08-165	11	Spring	2008	12.9	5.92	31.9
13	NFM08-168	14	Spring	2008	16.7	7.72	31.9
14	NFM08-169	2	Summer	2008	18.8	3.42	29.6
15	NFM08-170	3	Summer	2008	19.2	4.12	28.6
16	NFM08-173	6	Summer	2008	19.2	5.78	29.8
17	NFM08-175	8	Summer	2008	20.8	5.07	30.2
18	NFM08-178	11	Summer	2008	10.4	4.29	31.9
19	NFM08-180	14	Summer	2008	20.5	5.81	31.5

	A	B	C	D
1	collection_id	site	species	length_sl_mm
2	NFM08-142	2	Lagodon rhomboides	14.29
3	NFM08-142	2	Lagodon rhomboides	14.7
4	NFM08-142	2	Lagodon rhomboides	15.35
5	NFM08-142	2	Lagodon rhomboides	16.15
6	NFM08-142	2	Lagodon rhomboides	16.99
7	NFM08-142	2	Lagodon rhomboides	17.19
8	NFM08-142	2	Lagodon rhomboides	17.43
9	NFM08-142	2	Lagodon rhomboides	20.21
10	NFM08-142	2	Lagodon rhomboides	21.53
11	NFM08-142	2	Leiostomus xanthurus	18.63
12	NFM08-142	2	Leiostomus xanthurus	20.4
13	NFM08-143	3	Brevoortia patronus	23.58
14	NFM08-143	3	Fundulus similis	31.44
15	NFM08-143	3	Fundulus similis	50.22
16	NFM08-143	3	Fundulus similis	52.9
17	NFM08-143	3	Fundulus similis	54.7
18	NFM08-143	3	Fundulus similis	66.22
19	NFM08-143	3	Fundulus similis	69.19
20	NFM08-143	3	Lagodon rhomboides	14.66
21	NFM08-143	3	Lagodon rhomboides	16.77
22	NFM08-143	3	Lagodon rhomboides	18.61

Need several skills to summarize data or prepare for further analyses

habitat_type	season	n_species	n_individuals
erosional edge	Fall	7	370
erosional edge	Spring	8	664
erosional edge	Summer	8	80
erosional edge	Winter	7	4754
seagrass	Fall	9	74
seagrass	Spring	7	670
seagrass	Summer	6	102
seagrass	Winter	7	1508

Key packages and functions

dplyr

- left_join()/right_join()
- full_join()
- group_by()
- summarize()

- Join data frames based on common identifiers
- Group by and summarize

Pivot data to a wider format

- For, e.g., nMDS

	collection_id	Lagodon rhomboides	Leiostomus xanthurus	Brevoortia patronus	Fundulus similis	Mugil cephalus	Cyprinodon variegatus	
1	NFM08-142	9	2	0	0	0	0	
2	NFM08-143	3	0	1	6	0	0	
3	NFM08-146	56	80	122	0	843	0	
4	NFM08-148	12	16	6	0	0	0	
5	NFM08-151	1	12	7	0	4	3	
6	NFM08-154	3	1066	4	0	1	0	
7	NFM08-156	209	0	0	0	0	0	
8	NFM08-157	0	9	0	0	0	0	
9	NFM08-160	0	0	0	10	10	0	
10	NFM08-162	7	0	0	0	0	0	

Key packages and functions

tidyr
• `pivot_wider()`

C. elegans toxicity data

1 file: Change the shape; make it tidy

	A	B	C	D	E
1	Name	CASRN	Chemical.ID.number	C.elegans_response_0.5_uM	C.elegans_res
2	Cyanamide	420-04-2	34490	5.907811788	
3	Urea	57-13-6	21426	5.826716379	
4	Ethanolamine	141-43-5	22000	5.860733759	
5	Boric acid	10043-35-	20194	6.060643329	
6	Ethylene glycol	107-21-1	20597	5.790527793	
7					

Key packages and functions

tidyr

- pivot_longer()
- separate()

	Name	CASRN	Chemical.ID.number	Variable	Value
1	Cyanamide	420-04-2	34490	C.elegans_response_0.5_uM	5.907812
2	Cyanamide	420-04-2	34490	C.elegans_response_1.0_uM	5.955980
3	Cyanamide	420-04-2	34490	C.elegans_response_5.0_uM	6.012674
4	Cyanamide	420-04-2	34490	C.elegans_response_10_uM	5.998145
5	Cyanamide	420-04-2	34490	C.elegans_response_50_uM	5.963630
6	Cyanamide	420-04-2	34490	C.elegans_response_100_uM	6.014989
7	Cyanamide	420-04-2	34490	C.elegans_response_200_uM	5.920054
8	Urea	57-13-6	21426	C.elegans_response_0.5_uM	5.826716
9	Urea	57-13-6	21426	C.elegans_response_1.0_uM	5.860734
10	Urea	57-13-6	21426	C.elegans_response_5.0_uM	6.060643

	Name	CASRN	Chemical.ID.number	organism	parameter	dose	units	Value
1	Cyanamide	420-04-2	34490	C.elegans	response	0.5	uM	5.907812
2	Cyanamide	420-04-2	34490	C.elegans	response	1.0	uM	5.955980
3	Cyanamide	420-04-2	34490	C.elegans	response	5.0	uM	6.012674
4	Cyanamide	420-04-2	34490	C.elegans	response	10	uM	5.998145
5	Cyanamide	420-04-2	34490	C.elegans	response	50	uM	5.963630
6	Cyanamide	420-04-2	34490	C.elegans	response	100	uM	6.014989
7	Cyanamide	420-04-2	34490	C.elegans	response	200	uM	5.920054

Realistic expectations

What I hope you get out of this workshop

- Go write all your code from scratch, from memory

- Know **what's possible**
- **Identify** what you need to do
- Know enough **vocabulary** and help-finding techniques to **look up** what you need
- **Recognize solutions** when you find them
- Have enough **context** – and **confidence** – to implement and modify the solutions you find

Set yourself up for **success the next time** you open R

- RStudio projects
- Good file naming
- Code you understand: comment, comment, comment!
- Write easy-to-modify code
 - Example: If you have set up a report to summarize data from 2021, you may use 'year = 2021' in a lot of places. Subsetting, summarizing, etc.
 - Near the top of your script, make a variable you call 'my_year' and set it to 2021: 'my_year <- 2021'
 - Everywhere you see 'year = 2021' in your script, change that to 'year = my_year'
 - Next year, update that variable, in one place: 'my_year <- 2022'
- Don't put too much pressure on yourself: you will improve every time you update your code

Topics we will tackle

Really dissect and change a data frame

- Identify rows and columns:
 - By position
 - By characteristic
 - Defining TRUE/FALSE operations
- Subset a data frame (choose certain rows and/or columns)
 - Base R: `[]`
 - tidyverse: `dplyr`'s `select()` and `filter()`
- Modify a data frame [again, both base and tidyverse options]
 - Add variables/columns: `dplyr`'s `mutate()`
 - Perform calculations using columns
 - Change column names

Understanding different object types in R

- Data frames!
- Vectors
- Lists: especially useful for exploring statistical outputs

Do some coding: EDA

https://padlet.com/kim_cressman/oyh9vm6xugaxlk1p

Exploratory Data Analysis (EDA)

- Read in one (or more) of the discussed fish (lengths or wq) or veg files. Don't worry about joining data frames together; just work with them as-is.
- Figure out what's going on in the data frame: what are the columns? What are the rows? What class is each column (number, character, logical)?
- Make an exploratory graph (or 2, or 5....) and paste it in the appropriate column of the padlet. In your post, please answer these questions too:
 1. What variables did you put on the x- and y-axes?
 2. Did you do any grouping (for color, shape, facets, etc.)? What variables did you use for that?
 3. What summary statistics would you like to generate for this data frame?