

INFO101: Tabular Data

What makes data tidy?

Key concepts

Don't confuse the computer
Make it a rectangle

Make it a rectangle

	A	B	C
1	site	species	count
2	Santa Rosa	blue	3
3	Santa Rosa	fin	4
4	Santa Rosa	humpback	2
5	San Miguel	blue	4
6	San Miguel	fin	6
7	San Miguel	humpback	4
8	Santa Cruz	blue	5
9	Santa Cruz	fin	10
10	Santa Cruz	humpback	9

Non-rectangular examples

	A	B	C	D	E
1			species		
2			blues	fins	humpbacks
3		Santa Rosa	3	4	5
4	sites	San Miguel	4	6	10
5		Santa Cruz	2	4	9

Multiple headers, too confusing

	A	B	C	D
1	site	blues	fins	humpbacks
2	Santa Rosa	3	4	5
3	San Miguel	4	6	10
4	Santa Cruz	2	4	9

Don't confuse the computer

	A	B	C
1	latitude	depth_m	temp_c
2	45	5	10.6
3	45	100	7.1
4	30	5	21.8
5	30	100	18.3
6	15	5	27.1
7	15	100	22.6

Same variable types throughout = good

Confusing examples

latitude	depth	temp (°C)
45	5m	10.6
45	100m	7.1
30	5m	21.8
30	100m	18.3
15	5m	27.1
15	100m	22.6

latitude	5m	100m
45	10.6	7.1
30	21.8	18.3
15	27.1	22.6

Consistent names and formats

	A	B	C
1	date	air_temp_c	water_temp_c
2	2024-03-01	14.1	10.3
3	2024-03-02	NA	NA
4	2024-03-03	16.3	11.5
5	2024-03-04	17.8	11.2

Dates should follow universal
Headers should be computer-friendly

Missing values = NA

Inconsistent examples

date	air_temp_c	waterTempC
3/1/24	14.1	10.3
3/2/24	No survey	-
Mar 3 24	16.3	11.5
2024-03-04	17.8	11.2

Holy Heck, this hurts my eyes

Recap

Write good code!

Make it legible to humans and computers alike

New vocabulary and lingering questions

New vocabulary

Lingering questions

Exercises

Match the tables to the tidy rule they violate

l1	l2	b	c
-124.2	40.8	1	0
-124.3	40.7	1	0
-124.4	40.6	1	11
-124.5	40.5	2	0

Rule 3

Rule 1 - make it a rectangle

location	beaufort_state	count
-124.2, 40.8	1	0
-124.3, 40.7	1	0
-124.4, 40.6	1	11
-124.5, 40.5	2	0

Rule 2

Rule 2 - don't confuse the computer

# Marbled Murrelet at-sea survey data May 2015			
# Data collected by AJR, WEP, and LSI			
lon	lat	beaufort_state	count
-124.2	40.8	1	0
-124.3	40.7	1	0
-124.4	40.6	1	11
-124.5	40.5	2	0

Rule 1

Rule 3 - use consistent names and formats

INFO101: Tabular Data

Creating and importing data frames in R

MARINCS 100B | Intro to Marine Data Science | Winter 2025

Key concepts

Import dataframes

DFs are 2-D

Two views, same data

latitude	depth_m	temp_c
45	5	10.6
45	100	7.1
30	5	21.8
30	100	18.3

Creating a data frame

```
data_table <- data.frame(var1 = c(1, 2, 3, 4), var2 = c(5, 6, 7, 8), var3 = c(9, 10, 11, 12))
```

```
data_table <- read.csv("Data_table.csv")
```

Demo in R

Insert cool code shenanigans here

New vocabulary and lingering questions

New vocabulary

View() function is handy

colnames() is also good

Lingering questions

Exercises

Complete the exercises in `exercises/exercises101b.R`

INFO101: Tabular Data

Indexing data frames

MARINCS 100B | Intro to Marine Data Science | Winter 2025

Key concepts

index using []

index using row/column

How to index into data frames

latitude	depth_m	temp_c
45	5	10.6
45	100	7.1
30	5	21.8
30	100	18.3

latitude	depth_m	temp_c
1,1	1,2	1,3
2,1	2,2	2,3
3,1	3,2	3,3
4,1	4,2	4,3

`noaa_survey[1, 2]`

Pull rows and columns from data frames

latitude	depth_m	temp_c
45	5	10.6
45	100	7.1
30	5	21.8
30	100	18.3

`noaa_survey$latitude`
returns 1st column

`noaa_survey[1,]`

`noaa_survey[,1]`

Filtering rows

latitude	depth_m	temp_c
45	5	10.6
45	100	7.1
30	5	21.8
30	100	18.3

```
noaa_survey[noaa_survey$latitude == 45, ]
```

New vocabulary and lingering questions

New vocabulary

"buck"

filter by row/column

Lingering questions

Exercises

Complete the exercises in `exercises/exercises101c.R`