Ingesting object data from a BIO excel file

Dan Kelley

2021-05-28

Introduction

This directory contains an R file here in which I tried to decode things automatically. This proved to be difficult, because the meanings of things were (I think) indicated by colour underlays, and I don't know how to catch that nuance in a program. Also, there were *lots* of weird things in the file, like text in numeric fields, negative lengths, etc., and so the R file became a mess of special cases. So, and partly because this work would not pay off for other datasets, I decided to drop the idea of decoding the excel file with R, instead doing the work "by hand" by copying cells into csv files. (In addition to copying cells, I had to do computations, e.g. converting Newtons to kg.)

For the names of items, I switched from all upper-case to all lower-case, which is easier to type, and fits with the Dewey names. I also removed the `"` symbols, which are hard to type in code, and appear in some instances and not in others. And I change groups of multiple spaces into single spaces.

Details

`chains_bio.csv`

These are green-background in the excel file, rows 7 through 12.

`buoyancy`

This is inferred as `NEWTONS` (col E) divided by 9.8 m/s^2 to get kg.

```
N <- c(9.3408,1.46784,0.3336,0.40032,42.3672,445)
round(N/9.8, 3)
```

```
## [1]  0.953  0.150  0.034  0.041  4.323 45.408
```

`height`

The is copied directly from col B, with 0 for the first 4 entries.

`width`

This is inferred as the ratio of `A (m*2/m)` (col G) to `LENGTH` (col C)

```
G <- c(NA,NA,NA,NA,0.05715,0.0254)
C <- c(NA,NA,NA,NA,1,1)
length <- round(G/C, 4)
length
```

```
## [1]     NA     NA     NA     NA 0.0572 0.0254
```

`CD`

The `CD` is assigned the value 1.3, matching the Dewey assumption for chain elements.

Results

These are in the order of the spreadsheet; note that `chains("?")` merges all datasets, and orders by name.

```
knitr::kable(read.csv("chains_bio.csv"))
```

name	buoyancy	height	width	CD	code	source
swivel	-0.953	0	0.0000	1.3	NA	BIO
7/16in shackle	-0.150	0	0.0000	1.3	NA	BIO
1/4in nicopress sleeve	-0.034	0	0.0000	1.3	NA	BIO
5/16in heavy ss thimbles	-0.041	0	0.0000	1.3	NA	BIO
5/8in galvanized chain	-4.323	1	0.0572	1.3	NA	BIO
ballast weight	-45.408	1	0.0254	1.3	NA	BIO

`floats_bio.csv`

These are yellow-background in the excel file.

The first item, named `"new glass streamlined float"`, has no listed weight in pounds or newtons, so I ignored that entry.

For buoyancy, I took `NEWTONS` (col E) and divided by 9.8m/s^2 to get kg, rounding to 2 digits after the decimal place.

```
N <- c(445,285,560,516,429,773,414,1886,3963,4030,2095,8224,7740)
round(N/9.8, 2)
```

```
##  [1]  45.41  29.08  57.14  52.65  43.78  78.88  42.24 192.45 404.39 411.22
## [11] 213.78 839.18 789.80
```

There are two length columns in the file, but the values are either identical or relatable by rounding, so I could choose either. But is that in the x or the y direction?

Two columns contain things related to area, one with name `A (m*2/m)` and other with name `AW (m*2)`. I think the former is the projected area per meter of z (as it is for wires). So, if we assume that the length is in the z direction, or that the object is roughly spherical, we can infer what the package needs, namely `height` and `diameter`, from the ratio of the second area to the first length. Whether that's sensible, I just don't know.

```
A<-c(0.164,0.5,0.073,0.5,0.164,1.5,1.5,1.86,2.84,2.474,1.67,2.7,2.7)
L<-c(0.74,0.563,1,0.563,1.187,1,1,1.286,1.286,1.286,1.286,2.42,2.42)
round(A/L,3)
```

```
##  [1] 0.222 0.888 0.073 0.888 0.138 1.500 1.500 1.446 2.208 1.924 1.299 1.116
## [13] 1.116
```

```
f <- read.csv("floats_bio.csv")
o <- order(f$name)
knitr::kable(f[o,])
```

	name	buoyancy	height	diameter	CD	code	source
3	3 pack Viny 12b-3 floats	57.14	1.000	0.073	0.65	NA	BIO
2	a2 package adcp and 2 Viny balls	29.08	0.563	0.888	0.65	NA	BIO
7	ADCP / 2x c3 subs assembly	42.24	1.000	1.500	0.65	NA	BIO

	name	buoyancy	height	diameter	CD	code	source
5	bub 2x17in glass	43.78	1.187	0.138	0.65	NA	BIO
6	IPS / 2x b3 subs assembly	78.88	1.000	1.500	0.65	NA	BIO
9	stablemoor 1000lb 3500 msw with ADCP	404.39	1.286	2.208	0.65	NA	BIO
10	stablemoor 1015lb 1500 msw with ADCP	411.22	1.286	1.924	0.65	NA	BIO
8	stablemoor 533lb 3500 msw with ADCP	192.45	1.286	1.446	0.65	NA	BIO
11	stablemoor 580lb 1500 msw with ADCP	213.78	1.286	1.299	0.65	NA	BIO
1	streamlined bub 2 x 17in glass	45.41	0.740	0.222	0.65	NA	BIO
4	streamlined bub 3 Viny balls	52.65	0.563	0.888	0.65	NA	BIO
13	syn. float,bracket and 109lb ADCP	789.80	2.420	1.116	0.65	NA	BIO
12	syntactic float with ADCP bracket	839.18	2.420	1.116	0.65	NA	BIO

`instruments_bio.csv`

Processing

The entry for `DVS CURRENT METER` had height listed as `0.7 (2)`. The parenthetic value is not explained, and so it is ignored here, perhaps incurring an error.

The entry for `SBE37 MICROCAT CLAMP-ON STYLE*` (row 33) lists the value `(15)*` for Newtons, and the value 15N is used here, based on my interpretation of the footnote.

All names are converted to lower-case, and double spaces are converted to single spaces.

The entries in rows 34 and 35 seem to refer to one item. I ignore row 34, since so many things are not filled in for it, and I do not understand what the meaning is.

The `height` value is copied from the first `LENGTH` column (col C).

The `buoyancy` value is inferred by dividing the `NEWTONS` (col E) by 9.8m/s^2.

```
E <- c(73.392,106.752,131.216,177.92,44.48,200.16,444.8,209.056,199,192,21.8,45.6,600.48,15,144,216)
round(E/9.8, 2)
```

```
##  [1]  7.49 10.89 13.39 18.16  4.54 20.42 45.39 21.33 20.31 19.59  2.22  4.65
## [13] 61.27  1.53 14.69 22.04
```

The `area` is computed as the product of `LENGTH` (col C) and `A (m&2/m)` (col G).

The `CD` is assigned the value 0.65, matching the Dewey assumption for floats.

A `code` entry is made, to make this compatible with Dewey values, if I can find them.

Results

```
knitr::kable(read.csv("instruments_bio.csv"))
```

name	buoyancy	height	area	CD	code	source
seacat 16-03 (bar,plastic case)	-7.49	0.8400	0.122	0.65	NA	BIO
seacat 16-04 (bar,titanium no press.)	-10.89	0.8763	0.062	0.65	NA	BIO
seacat 16-04 (bar,titanium with press.)	-13.39	1.0668	0.077	0.65	NA	BIO
seabird CTD (ios oxygen with bar)	-18.16	1.2700	0.094	0.65	NA	BIO
DVS current meter	-4.54	0.7000	0.243	0.65	NA	BIO
wotan (bar)	-20.42	1.6800	0.176	0.65	NA	BIO
RDI with bar	-45.39	2.1800	0.270	0.65	NA	BIO
RCM-8 with fin	-21.33	0.7400	0.060	0.65	NA	BIO
RCM-11 in frame	-20.31	0.8000	0.078	0.65	NA	BIO

name	buoyancy	height	area	CD	code	source
seaguard (2000 & 6000 m)	-19.59	0.8600	0.083	0.65	NA	BIO
vemco on a ss bar	-2.22	1.2000	0.052	0.65	NA	BIO
SBE37 microcat on a ss bar	-4.65	1.2000	0.060	0.65	NA	BIO
sediment trap (#1349)	-61.27	1.8290	1.061	0.65	NA	BIO
SBE37 microcat clamp-on style	-1.53	0.5590	0.117	0.65	NA	BIO
WH ADCP sentinel (orange) 500m with inline frame	-14.69	0.7600	0.091	0.65	NA	BIO
WH ADCP sentinel (yellow) 6000m with inline frame	-22.04	0.7900	0.111	0.65	NA	BIO

`releases_bio.csv`

Note the ommision of an entry for type `BENTHOS 875 RELEASE`, because it has no value listed in col G (area `A (m*2/m)`) or col I (area `AW (m*2)`).

`height`

Use `LENGTH` (col C).

`buoyancy`

Use the negative of `NEWTONS` (col E), divided by g=9.8m/s^2, rounded to 2 decimal places.

```
N <- c(149, 253.5, 360, 129, 273, 118.7)
round(N/9.8, 2)
```

```
## [1] 15.20 25.87 36.73 13.16 27.86 12.11
```

`width`

Use the ratio of `A (m&2/m)` (col G) to `LENGTH` (col C).

```
G <- c(0.14,0.14,0.127,0.112,0.199,0.237)
C <- c(1.194,1.32,1.23,0.68,1,0.4)
width <- round(G/C, 4)
```

Note the `NA` value – this is "benthos 875 release", which we will not use here, since there's no point in guessing. (Put another way, if someone at BIO wants to take a measuring tape to that thing to find it's diameter, we'll incorporate it!)

`CD`

Use 0.65, i.e. Dewey's value for floats, on the assumption that the two might be similar.

Result for releases

```
knitr::kable(read.csv("releases_bio.csv"))
```

name	buoyancy	height	width	CD	code	source
EG&G 723a	-15.20	1.194	0.1173	0.65	NA	BIO
EG&G 723a with tension bar	-25.87	1.320	0.1061	0.65	NA	BIO
benthos 965a release	-36.73	1.230	0.1033	0.65	NA	BIO
benthos 966a and 866a release	-13.16	0.680	0.1647	0.65	NA	BIO
benthos 865a release	-27.86	1.000	0.1990	0.65	NA	BIO

name	buoyancy	height	width	CD	code	source
vr2w release with float collar	-12.11	0.400	0.5925	0.65	NA	BIO

wires_bio.csv

`buoyancyPerMeter`

The 3/16 case has -0.76 in a column named "W Nt/m", so I've converted that to -0.76/9.8=-0.0776 (rounded to 4 places), to get kg/m instead of N/m.

There is no listing for the 1/4 case, but I estimated (guessed) that as

```
round(-0.0776 * ((4/16)/(3/16))^2, 4)
```

```
## [1] -0.138
```

This is based on an assumption that the metal inside is similar, etc.

`diameter`

For the diameter, I looked at the column labeled `A (m*2/m)`. I assume that is the horizontally-projected area. To test that, I computed

```
16*c(0.00635,0.0079375)/0.0254
```

```
## [1] 4 5
```

which yields the expected, for the outside diameters in 16-ths of inch.

`CD`

There are no entries in the file and I did not know what `DRAG -N` and `DRAG -T` meant and so I just used the value $C_D = 1.3$, used in Dewey's work.

Result for `wires`

```
knitr::kable(read.csv("wires_bio.csv"))
```

name	buoyancyPerMeter	diameter	CD	code	source
3/16in galvanized wire coated to 1/4in	-0.0775	0.0063500	1.3	NA	BIO
1/4in galvanized wire coated to 5/16in	-0.1380	0.0079375	1.3	NA	BIO