Ingesting object data from a BIO excel file

Dan Kelley

2021-05-28, 2021-06-08, 2021-06-09 and 2021-06-25

Contents

1 Document Revisions

1. 2021-06-25
    1. Floats store area, not diameter
    2. Connectors store area, not width
    3. Wires and chains store areaPerMeter instead of diameter or width
2. 2021-06-08 (thanks to @formulafunracing and @clayton33 for advice on this)
    1. Move discrete chain items like shackles to a new entity, "connector".
    2. Make `chain` objects have `length`, not `height`.
    3. Rename some items for consistency and readability.
    4. Add `originalName` column.

2 Introduction

This directory contains a file named `bio.R` in which I tried to decode things automatically. This proved to be difficult, because the meanings of things are (I gather) indicated by colour underlays, and I don't know how to catch that nuance in a program. Another problem was that there are *lots* of weird things in the excel file, like text in numeric fields, negative lengths, etc., and so the R code for renaming became a mess of special cases. After a while, I just gave up on decoding the excel file with R code, instead creating csv files semi-manually, by copy-pasting into this file to do some calculations, then copy-pasting back to csv. I outline some of the steps here but, as is always the case for non-automatic work, this record is best regarded as a rough sketch.

The names were altered in an attempt to improve consistency within these datasets and also with the Dewey datasets. In this second category, for example, I removed double spaces, and replaced `"` characters with the string `in`, for inches. The original names are stored in the column named `originalNames` and are verbatim from the excel file, *except* that multiple consecutive spaces are converted to single spaces, leading and trailing spaces are removed, and trailing asterisks (for footnotes) are removed.

3 Details

3.1 chains_bio.csv and connectors_bio.csv

These are green-background in the excel file, rows 7 through 12. I put `5/8" galvanized chain` (row 11) into `chains_bio.csv` and the rest into `connectors_bio.csv`.

No lengths or areas are listed for the first 4 items, but height is set to 0.05m here, as a plausibe estimate. Similarly, the area is set to 0.0025m^2, on the assumption that the item is 5cm wide and 5cm tall. These are both such small values that they will have little effect on the calculations.

3.1.1 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
```

3.1.2 height

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

3.1.3 area (replaces 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)
(area <- round(G*C, 4))
```

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

3.1.4 CD

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

3.1.5 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"), caption="Contents of chains_bio.csv")
```

Table 1: Contents of chains_bio.csv

name	buoyancyPerMeter	areaPerMeter	CD	code	source	originalName
5/8in galvanized chain	-4.323	0.0572	1.3	NA	BIO	5/8" GALVANIZED CHAIN

and here are the connectors:

```r
knitr::kable(read.csv("connectors_bio.csv"), caption="Contents of connectors_bio.csv")
```

Table 2: Contents of connectors_bio.csv

name	buoyancy	height	area	CD	code	source	originalName
swivel	-0.953	0.05	0.0025	1.3	NA	BIO	SWIVEL
7/16in shackle	-0.150	0.05	0.0025	1.3	NA	BIO	7/16" SHACKLE
1/4in nicopress sleeve	-0.034	0.05	0.0025	1.3	NA	BIO	1/4" NICOPRESS SLEEVE
5/16in heavy SS thimbles	-0.041	0.05	0.0025	1.3	NA	BIO	5/16" HEAVY SS THIMBLES
ballast weight	-45.408	1.00	0.0254	1.3	NA	BIO	BALLAST WEIGHT

3.2 `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, but a personal communication suggests using buoyancy 100 pounds, i.e. 445N i.e. 45.41kg.

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.

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

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

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, colG labelled `A (m*2/m)` and colI labelled `AW (m*2)`. From https://github.com/dankelley/mooring/issues/24#issuecomment-857715929, we know that the frontal area is given by the product of colC and colG.

We take as a test case: the `"BUB 2x178" GLASS"` value, in row 42. The above-stated URL indicates that a BUB is a package with two 17in glass spheres, with height (length) 1.187m (which matches colC) and frontal area 0.5531 m^2. As a check, the product of colC (1.187) and colG (0.466) is 0.553142m^2, which matches the value stated on the above-stated URL.

So, area will be computed as the product of colC and colG.

```r
# Notice that row 37 is ignored, because no weight is listed for it
colC <- c(0.87, 0.74,0.563,1,0.563,1.187,1,1,1.286,1.286,1.286,1.286,2.42,2.42)
colG <- c(0.21, 0.22,0.22,0.22,0.22,0.466,0.44,0.65,1.822,1.822,1.807,1.822,0.51,0.51)
area <- colG*colC
round(area, 3)
```

```
##  [1] 0.183 0.163 0.124 0.220 0.124 0.553 0.440 0.650 2.343 2.343 2.324 2.343
## [13] 1.234 1.234
```

```r
f <- read.csv("floats_bio.csv")
knitr::kable(f, caption="Contents of floats_bio.csv")
```

Table 3: Contents of floats_bio.csv

name	buoyancy	height	area	CD	code	source	originalName
new glass streamlined float c2	45.41	0.870	0.183	0.6	NA	BIO	NEW GLASS STREAMLINED FLOAT (C2)
streamlined BUB 2x 17in glass	45.41	0.740	0.163	0.6	NA	BIO	STREAMLINED BUB 2 x 17" GLASS
a2 package ADCP and 2 Viny balls	29.08	0.563	0.124	0.6	NA	BIO	A2 PACKAGE ADCP AND 2 VINY BALLS
3 pack Viny 12b-3 floats	57.14	1.000	0.220	1.1	NA	BIO	3 pack VINY 12B-3 floats
streamlined BUB 3 Viny balls	52.65	0.563	0.124	0.6	NA	BIO	STREAMLINED BUB 3 VINY BALLS
BUB 2x17in glass	43.78	1.187	0.553	1.1	NA	BIO	BUB 2x17" GLASS
IPS / 2x b3 subs assembly	78.88	1.000	0.440	0.6	NA	BIO	IPS / 2x B3 SUBS ASSEMBLY
ADCP / 2x c3 subs assembly	42.24	1.000	0.650	0.6	NA	BIO	ADCP / 2x C3 SUBS ASSEMBLY
stablemoor 533lb 3500 msw with ADCP	192.45	1.286	2.343	0.6	NA	BIO	Stablemoor 533 lb 3500 msw with ADCP
stablemoor 1000lb 3500 msw with ADCP	404.39	1.286	2.343	0.6	NA	BIO	Stablemoor 1000 lb 3500 msw with ADCP
stablemoor 1015lb 1500 msw with ADCP	411.22	1.286	2.324	0.6	NA	BIO	Stablemoor 1015 lb 1500 msw with ADCP
stablemoor 580lb 1500 msw with ADCP	213.78	1.286	2.343	0.6	NA	BIO	Stablemoor 580 lb 1500 msw with ADCP
syn. float with ADCP bracket	839.18	2.420	1.234	0.6	NA	BIO	SYNTACTIC FLOAT WITH ADCP BRACKET
syn. float, bracket and 109lb ADCP	789.80	2.420	1.234	0.6	NA	BIO	SYN. FLOAT, BRACKET AND 109 LB.ADCP

3.3 `instruments_bio.csv`

3.3.1 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.

3.3.2 Results

```
knitr::kable(read.csv("instruments_bio.csv"), caption="Contents of instruments_bio.csv")
```

Table 4: Contents of instruments_bio.csv

name	buoyancy	height	area	CD	code	source	originalName
seacat 16-03 (bar,plastic case)	-7.49	0.8400	0.122	0.65	NA	BIO	SEACAT 16-03 (BAR, PLASTIC CASE)
seacat 16-04 (bar,titanium no press.)	-10.89	0.8763	0.062	0.65	NA	BIO	SEACAT 16-04 (BAR,TITANIUM NO PRESS.)
seacat 16-04 (bar,titanium with press.)	-13.39	1.0668	0.077	0.65	NA	BIO	SEACAT 16-04 (BAR,TITANIUM WITH PRESS.)
seabird CTD (ios oxygen with bar)	-18.16	1.2700	0.094	0.65	NA	BIO	SEABIRD CTD (IOS OXYGEN WITH BAR)
DVS current meter	-4.54	0.7000	0.243	0.65	NA	BIO	DVS CURRENT METER
wotan (bar)	-20.42	1.6800	0.176	0.65	NA	BIO	WOTAN (BAR)
RDI with bar	-45.39	2.1800	0.270	0.65	NA	BIO	RDI WITH BAR
RCM-8 with fin	-21.33	0.7400	0.060	0.65	NA	BIO	RCM-8 WITH FIN
RCM-11 in frame	-20.31	0.8000	0.078	0.65	NA	BIO	RCM-11 IN FRAME
seaguard (2000 & 6000 m)	-19.59	0.8600	0.083	0.65	NA	BIO	SEAGUARD (2000 & 6000 M)
vemco on a ss bar	-2.22	1.2000	0.052	0.65	NA	BIO	VEMCO ON A SS BAR
SBE37 microcat on a ss bar	-4.65	1.2000	0.060	0.65	NA	BIO	SBE37 MICROCAT ON A SS BAR
sediment trap	-61.27	1.8290	1.061	0.65	NA	BIO	SEDIMENT TRAP (#1349)
SBE37 microcat clamp-on style	-1.53	0.5590	0.117	0.65	NA	BIO	SBE37 MICROCAT CLAMP-ON STYLE
ADCP sentinel (orange) 500m with inline frame	-14.69	0.7600	0.091	0.65	NA	BIO	WHADCP Sentinel (orange) 500 meter with inline frame
ADCP sentinel (yellow) 6000m with inline frame	-22.04	0.7900	0.111	0.65	NA	BIO	WHADCP Sentinel (yellow) 6000 meter with inline frame

3.4 `releases_bio.csv`

Notes: 1. The release data are in rows 13 to 19 in the spreadsheet (with grey underlay). 2. There is no 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)`). 3. The word 'release' is dropped from names, since it's redudant from the object class.

3.4.1 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
```

3.4.2 height

Use `LENGTH` (col C), ignoring row 16 (BENTHOS 875 RELEASE) since no area for that.

```
(C <- c(1.194, 1.32, 1.23,  0.68,  1,     0.4))
```

```
## [1] 1.194 1.320 1.230 0.680 1.000 0.400
```

3.4.3 area

Use the product of `LENGTH/m` (col C) and `A (m&2/m)` (col G) . Note that there is also a column labelled `LENGTH(m)` (col F) but this is ignored because (a) it has lots of blank values, and (b) where there *are* values, they appear to be no more than rounded versions col C. Also, note that there is no area reported for `BENTHOS 875 RELEASE`, so that value (row 16) is omitted.

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

```
## [1] 0.1672 0.1848 0.1562 0.0762 0.1990 0.0948
```

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!)

3.4.4 OLD: delete

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!)

3.4.5 CD

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

3.4.6 Result for releases

```
knitr::kable(read.csv("releases_bio.csv"), caption="Contents of releases_bio.csv")
```

Table 5: Contents of releases_bio.csv

name	buoyancy	height	area	CD	code	source	originalName
EG&G 723a	-15.20	1.194	0.1173	0.65	NA	BIO	EG&G 723A
EG&G 723a with tension bar	-25.87	1.320	0.1848	0.65	NA	BIO	EG&G 723A WITH TENSION BAR
benthos 965a	-36.73	1.230	0.1562	0.65	NA	BIO	BENTHOS 965A RELEASE
benthos 966a and 866a	-13.16	0.680	0.0762	0.65	NA	BIO	BENTHOS 966A and 866A RELEASE
benthos 865a	-27.86	1.000	0.1990	0.65	NA	BIO	BENTHOS 865A RELEASE
vr2w with float collar	-12.11	0.400	0.0948	0.65	NA	BIO	VR2W RELEASE with float collar

3.5 wires_bio.csv

3.5.1 `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.

3.5.2 `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.

3.5.3 `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.

3.5.4 Result for `wires`

```
knitr::kable(read.csv("wires_bio.csv"), caption="Contents of wires_bio.csv")
```

Table 6: Contents of wires_bio.csv

name	buoyancyPerMeter	areaPerMeter	CD	code	source	originalName
3/16in galvanized wire coated to 1/4in	-0.0775	0.0063500	1.3	NA	BIO	3/16" GALVANIZED WIRE COATED TO 1/4

name	buoyancyPerMeter	areaPerMeter	CD	code	source	originalName
1/4in galvanized wire coated to 5/16in	-0.1380	0.0079375	1.3	NA	BIO	1/4" GALVANIZED WIRE COATED TO 5/16