

epicollect field sample metadata

```
# Tidy epicollect data

library(here)

## here() starts at /home/cengstro/ownCloud/proj/radiometer
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.5          v purrr  0.3.4
## v tibble  3.1.5          v dplyr  1.0.7
## v tidyr   1.1.4          v stringr 1.4.0.9000
## v readr   2.0.2          v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(janitor)

##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##   chisq.test, fisher.test

library(lubridate)

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union

library(kableExtra)

##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##   group_rows

raw_epi <- read_csv(here("data/field_sample_meta/form-1__radiometer.csv"))

## Rows: 32 Columns: 17

## -- Column specification -----
## Delimiter: ","
## chr  (9): ec5_uuid, created_by, title, 1_ID, UTM_Zone_2_coords, 3_closeup_ph...
```

```

## dbl (6): lat_2_coords, long_2_coords, accuracy_2_coords, UTM_Northing_2_coo...
## dtm (2): created_at, uploaded_at

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
final_list <- read_csv(here("data/field_sample_meta/final_sample_list.csv"))

## Rows: 26 Columns: 2

## -- Column specification -----
## Delimiter: ","
## chr (2): sample_id, type

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
glimpse(raw_epi)

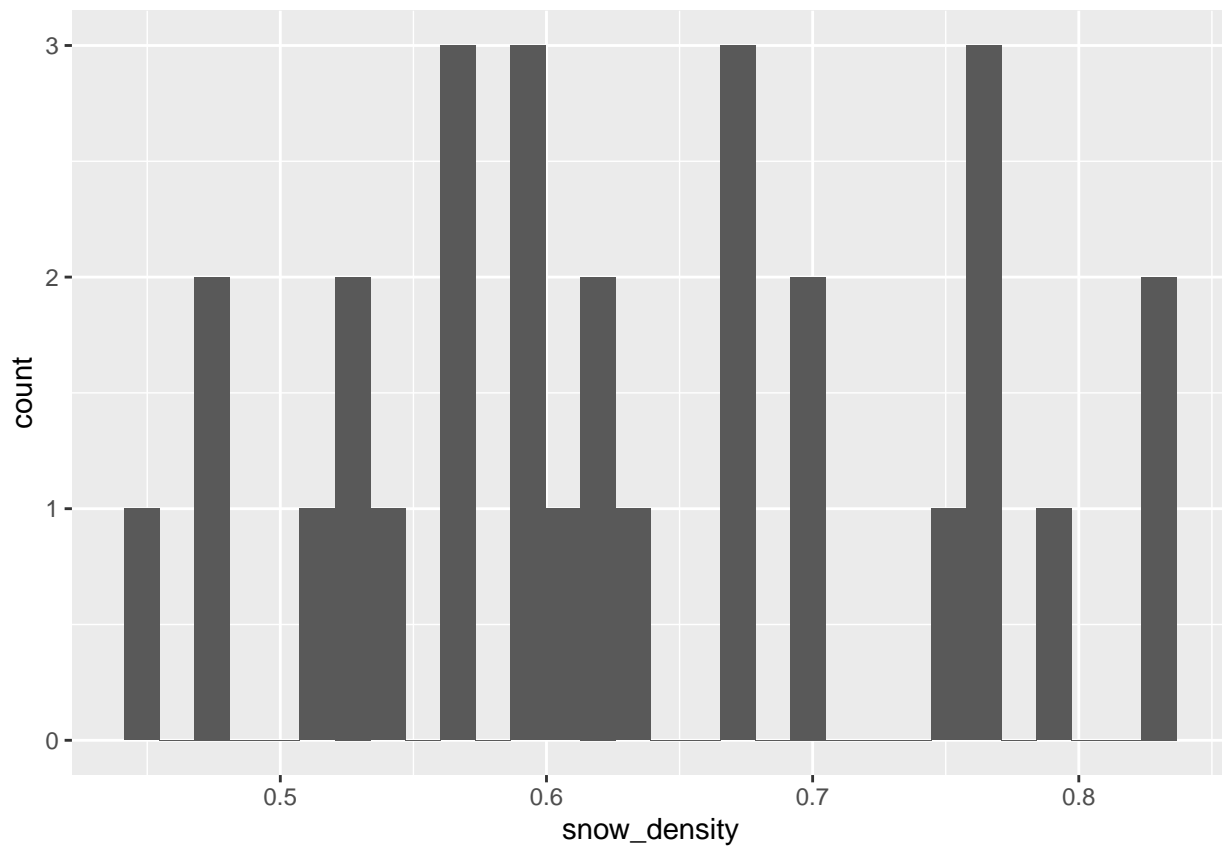
## Rows: 32
## Columns: 17
## $ ec5_uuid          <chr> "94f69fe4-dfc0-4610-a893-1f019b7e6918", "c2a99e~
## $ created_at        <dtm> 2021-08-05 21:46:12, 2021-08-05 21:30:29, 2021~
## $ uploaded_at       <dtm> 2021-08-06 01:35:29, 2021-08-06 01:35:27, 2021~
## $ created_by        <chr> "casey.engstrom@gmail.com", "casey.engstrom@gma~
## $ title             <chr> "Bdw21.11", "Bdw21.10", "Bdw21.09", "Bdw21.08",~
## $ `1_ID`           <chr> "Bdw21.11", "Bdw21.10", "Bdw21.09", "Bdw21.08",~
## $ lat_2_coords      <dbl> 50.10857, 50.10868, 50.10871, 50.10874, 50.1086~
## $ long_2_coords     <dbl> -123.1992, -123.1995, -123.1993, -123.1992, -12~
## $ accuracy_2_coords <dbl> 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 5, 4, 4, 4, 3, 6,~
## $ UTM_Northing_2_coords <dbl> 5550721, 5550733, 5550736, 5550740, 5550730, 55~
## $ UTM_Easting_2_coords <dbl> 485756, 485736, 485750, 485758, 485763, 485747,~
## $ UTM_Zone_2_coords  <chr> "10U", "10U", "10U", "10U", "10U", "10U", "10U"~
## $ `3_closeup_photo` <chr> NA, "c2a99e71-19fd-4b35-ae45-a6d418d06e8d_16281~
## $ `4_site_photo`    <chr> "94f69fe4-dfc0-4610-a893-1f019b7e6918_162819996~
## $ `5_notes`         <chr> NA, "Sed114-115", "Sed110-111", "Sed112-113", "~
## $ `6_algae_container_ty` <chr> "falcon", "ziploc", "ziploc", "ziploc", "ziploc~
## $ `7_wt_g`          <dbl> NA, 90, 117, 82, 120, 74, 93, 80, 125, 93, 85, ~

epi_select <- raw_epi %>%
  # rename for convenience
  rename(datetime = created_at,
    sample_id = `1_ID`,
    lat = lat_2_coords,
    lon = long_2_coords,
    scan_strings = `5_notes`, # the sed file number
    container_type = `6_algae_container_ty`,
    weight_g = `7_wt_g`) %>%
  # discard unused cols
  select(datetime, sample_id, lon, lat, scan_strings, container_type, weight_g) %>%
  # extract date
  mutate(date = as_date(datetime),
    sample_id = sample_id %>% tolower(),
    .after = sample_id)
epi_select

```

```
## # A tibble: 32 x 8
##   datetime      sample_id date      lon   lat scan_strings
##   <dtm>         <chr>   <date>   <dbl> <dbl> <chr>
## 1 2021-08-05 21:46:12 bdw21.11 2021-08-05 -123.  50.1 <NA>
## 2 2021-08-05 21:30:29 bdw21.10 2021-08-05 -123.  50.1 Sed114-115
## 3 2021-08-05 21:28:20 bdw21.09 2021-08-05 -123.  50.1 Sed110-111
## 4 2021-08-05 21:25:20 bdw21.08 2021-08-05 -123.  50.1 Sed112-113
## 5 2021-08-05 21:21:46 bdw21.07 2021-08-05 -123.  50.1 Sed108-109
## 6 2021-08-05 21:17:11 bdw21.06 2021-08-05 -123.  50.1 Sed106-107
## 7 2021-08-05 21:13:05 bdw21.04 2021-08-05 -123.  50.1 Sed104-105
## 8 2021-08-05 21:09:23 bdw21.03 2021-08-05 -123.  50.1 Sed102-103
## 9 2021-08-05 21:05:24 bdw21.02 2021-08-05 -123.  50.1 Sed100-101
## 10 2021-08-03 23:23:14 tri21.15 2021-08-03 -123.  50.0 Sed195-196
## # ... with 22 more rows, and 2 more variables: container_type <chr>,
## #   weight_g <dbl>
# what is the mean snow density? assuming  $\pi \times 5\text{cm}^2 \times 2\text{ cm volume} = 157\text{ ml}$ 
epi_select %>%
  mutate(snow_density = weight_g/157) %>% # in g/ml, see wiki firn snow density
  select(sample_id, snow_density) %>%
  ggplot(aes(snow_density)) +
  geom_histogram()

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 3 rows containing non-finite values (stat_bin).
```



```

# the tidy_radiometer data are grouped by scan_id, in the format 20210803_00001
# split each string into chr vector containing each scan_id that pertains to that sample

# # test
# "sed 111-113" %>%
#   str_extract_all("[:digit:]+") %>%
#   unlist() %>%
#   as.numeric() %>%
#   reduce(seq)
# # end test

get_vec_of_scan_ids <- function(str){
  str %>%
    str_extract_all("[:digit:]+") %>%
    unlist() %>%
    as.numeric() %>%
    reduce(seq) %>%
    map_chr(str_pad, 5, pad = "0")
}
# "sed 111-113" %>% get_vec_of_scan_ids() #test

# apply the function to the list
scan_str_list <- epi_select$scan_strings %>%
  as.list() %>%
  map(get_vec_of_scan_ids)

epi_scan_id <- epi_select %>%
  mutate(scan_strings = scan_str_list) %>% # replace the old with the new list, better way to do this?
  unnest_longer(scan_strings) %>%
  mutate(scan_id = str_glue("{date}_{scan_strings}") %>%
    str_remove_all("-"),
    .keep = "unused")

epi_scan_id %>%
  filter(!str_detect(scan_id, "NA")) %>%
  write_csv(here("data/field_sample_meta/tidy_epicollect.csv"))

# make a table of sample metadata

kk <- epi_select %>%
  mutate(sample_id = sample_id %>% str_remove("21.0"),
    sample_id = sample_id %>% str_remove("21.")) %>% # remove the 21 and unpad 0 for simplicity
  inner_join(final_list) %>%
  # convert to local time
  mutate(datetime = datetime - hours(8)) %>%
  select(sample_id, type, lat, lon, datetime)

## Joining, by = "sample_id"

kk %>%
  kbl(format = "latex", booktabs = TRUE) %>%
  kable_styling(latex_options = "striped")

```

sample_id	type	lat	lon	datetime
bdw10	white	50.10868	-123.1995	2021-08-05 13:30:29
bdw9	white	50.10871	-123.1993	2021-08-05 13:28:20
bdw8	red	50.10874	-123.1992	2021-08-05 13:25:20
bdw7	white	50.10866	-123.1991	2021-08-05 13:21:46
bdw6	red	50.10863	-123.1993	2021-08-05 13:17:11
bdw4	red	50.10867	-123.1991	2021-08-05 13:13:05
bdw3	red	50.10860	-123.1990	2021-08-05 13:09:23
bdw2	red	50.10856	-123.1989	2021-08-05 13:05:24
tri15	red	50.00730	-123.2376	2021-08-03 15:23:14
tri14	dirty	50.00738	-123.2375	2021-08-03 15:19:07
tri12	red	50.00745	-123.2376	2021-08-03 15:13:09
tri11	red	50.00745	-123.2377	2021-08-03 15:06:57
tri10	red	50.00744	-123.2378	2021-08-03 15:04:45
tri9	red	50.00744	-123.2378	2021-08-03 15:01:10
tri6	red	50.00784	-123.2406	2021-08-03 14:15:35
tri5	white	50.00780	-123.2406	2021-08-03 13:05:02
tri4	red	50.00781	-123.2407	2021-08-03 13:01:32
tri3	red	50.00794	-123.2406	2021-08-03 12:58:14
whi10	dirty	50.09127	-122.8777	2021-07-30 13:57:35
whi9	white	50.09107	-122.8776	2021-07-30 13:53:36
whi8	white	50.09110	-122.8776	2021-07-30 13:50:28
whi7	white	50.09116	-122.8774	2021-07-30 13:45:55
whi6	red	50.09117	-122.8773	2021-07-30 13:40:55
whi5	red	50.09116	-122.8772	2021-07-30 13:35:01
whi4	red	50.09129	-122.8772	2021-07-30 13:22:28
whi3	red	50.09127	-122.8772	2021-07-30 13:17:35