

Analysis of potential sample size

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.4
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.4.4      v tibble     3.2.1
v lubridate  1.9.3      v tidyr      1.3.0
v purrr      1.0.2
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
library(dbplyr)
```

Attaching package: 'dbplyr'

The following objects are masked from 'package:dplyr':

ident, sql

```
library(RSQLite)
```

Download FIA database locally

Only run once. Future - make a regular cron job to download the csv files that can be hosted on s3 (then arrow can be used to access)

```
download.file("https://apps.fs.usda.gov/fia/datamart/Databases/SQLite_FIADB_ENTIRE.zip",
             "SQLite_FIADB_ENTIRE.zip",
             method = "curl")
unzip("SQLite_FIADB_ENTIRE.zip")
```

Connect to database

```
fia_all <- DBI::dbConnect(RSQLite::SQLite(), "SQLite_FIADB_ENTIRE.db")
```

Access tree table

```
trees <- tbl(fia_all, "TREE")
```

Analyze sample size

Read in the table of focal species from Mike Bell

```
focal_species <- read_csv("focal_species.csv", show_col_types = FALSE)
```

Calculate the number of remeasurements for each unique tree in the database.

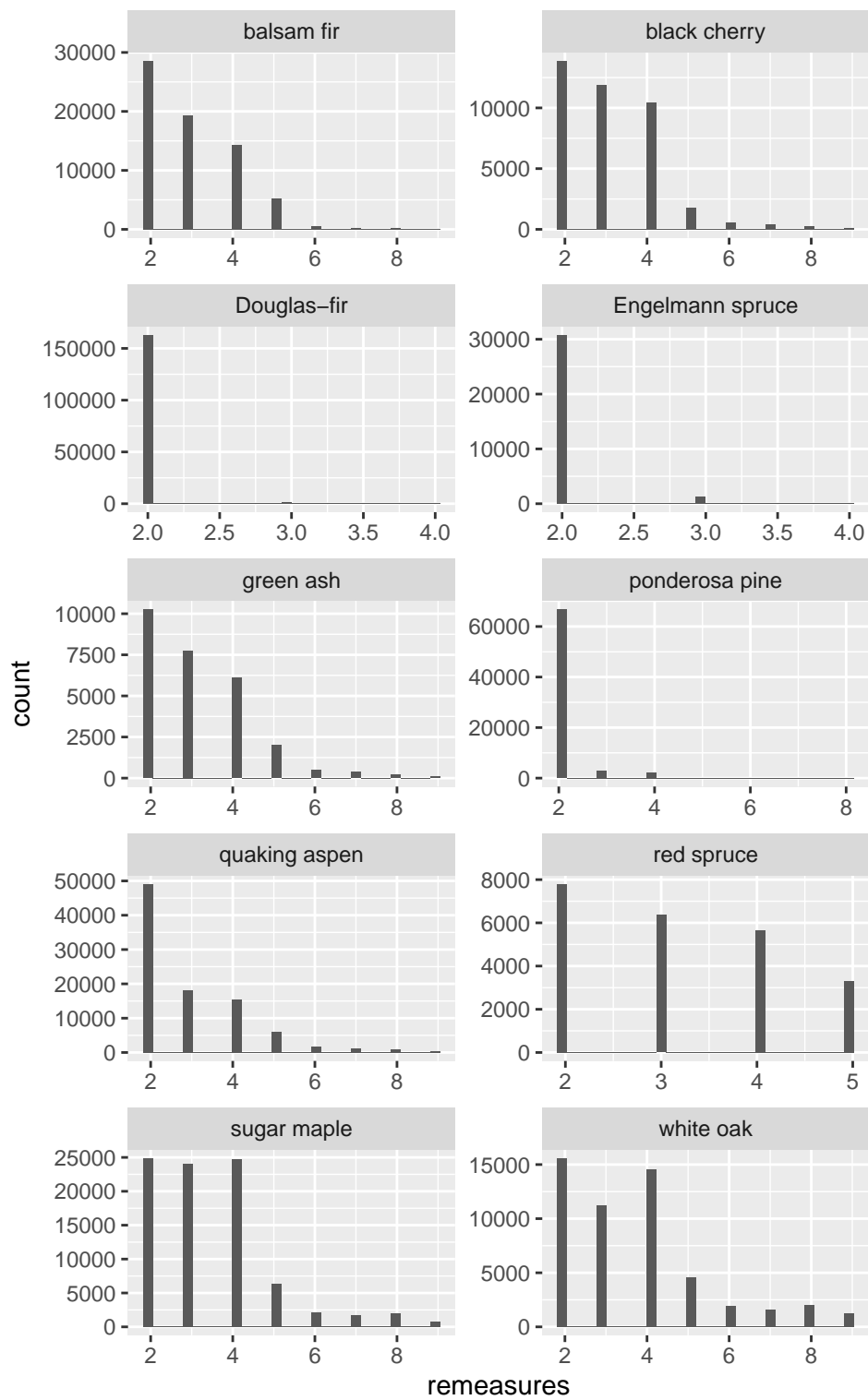
```
focal_SPCD <- focal_species$SPCD
tree_stats <- trees |>
  select(SUBP, TREE, PLOT, STATECD, SPCD, INVYR, DIA) |>
  filter(SPCD %in% focal_SPCD,
         INVYR > 1995,
         DIA > 5.0) |>
  mutate(id = paste(STATECD, PLOT, SUBP, TREE, sep = "-")) |>
  summarise(remeasures = n(), .by = c("id", "SPCD")) |>
  filter(remeasures > 1 & remeasures < 10) |>
  collect()
```

Remeasurements by species

This shows the max number of remeasurements per tree for the different species.

```
tree_stats |>
  left_join(focal_species, by = "SPCD") |>
  ggplot(aes(x = remeasures)) +
  geom_histogram() +
  facet_wrap(~common_name, scales = "free", ncol = 2)
```

``stat_bin()`` using ``bins = 30``. Pick better value with ``binwidth``.



Sample size by species

This table is the potential sample size per species. It is the number of trees that have been remeasured.

```
tree_stats |>
  summarise(sample_size = n(), .by = "SPCD") |>
  left_join(focal_species, by = "SPCD") |>
  select(common_name, sample_size) |>
  knitr::kable()
```

common_name	sample_size
white oak	52646
black cherry	39417
green ash	27271
sugar maple	86544
Douglas-fir	163998
quaking aspen	92940
Engelmann spruce	31978
ponderosa pine	72085
balsam fir	68643
red spruce	23109