Script to check accuracy and % of missing EUNIS for EVA and EVA Resurvey

Set global chunk options

```
knitr::opts_chunk$set(echo = FALSE)
```

Load libraries

Define functions

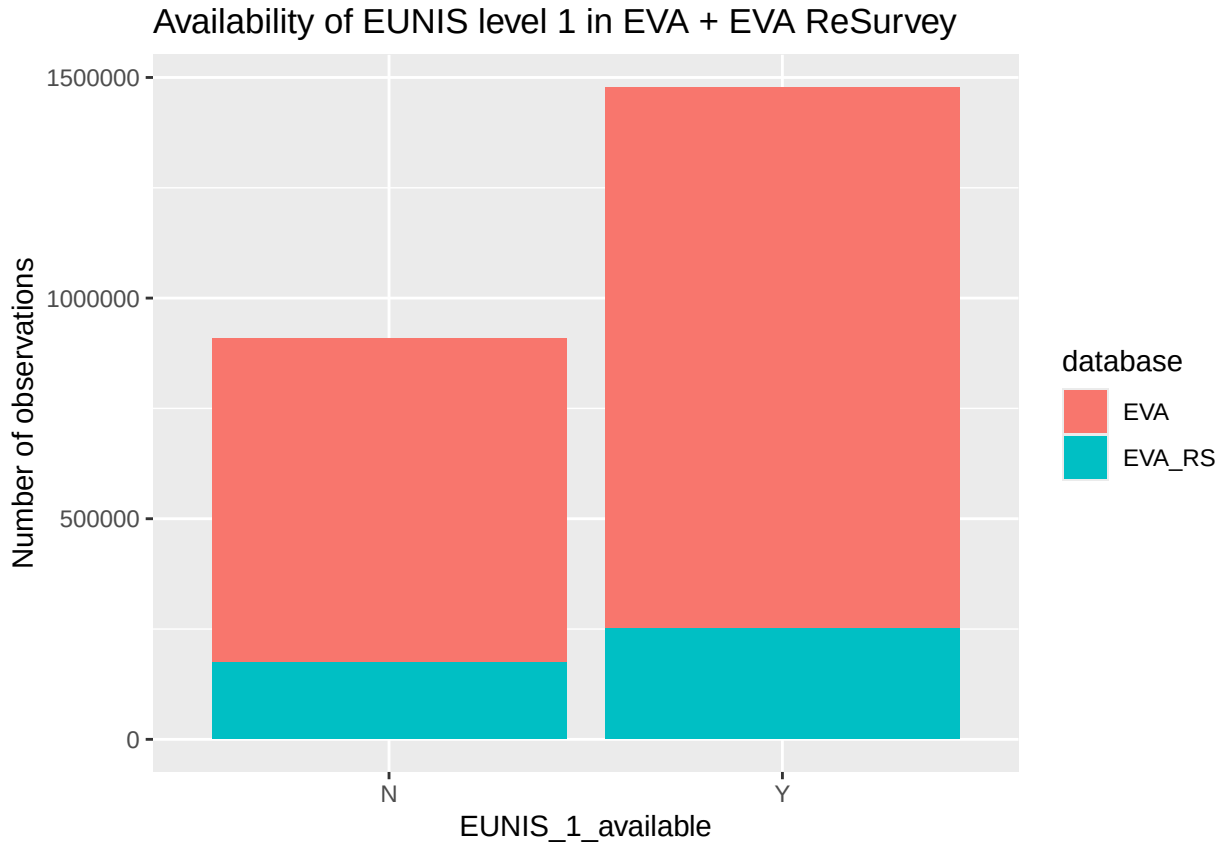
Read the data

No parsing issues!

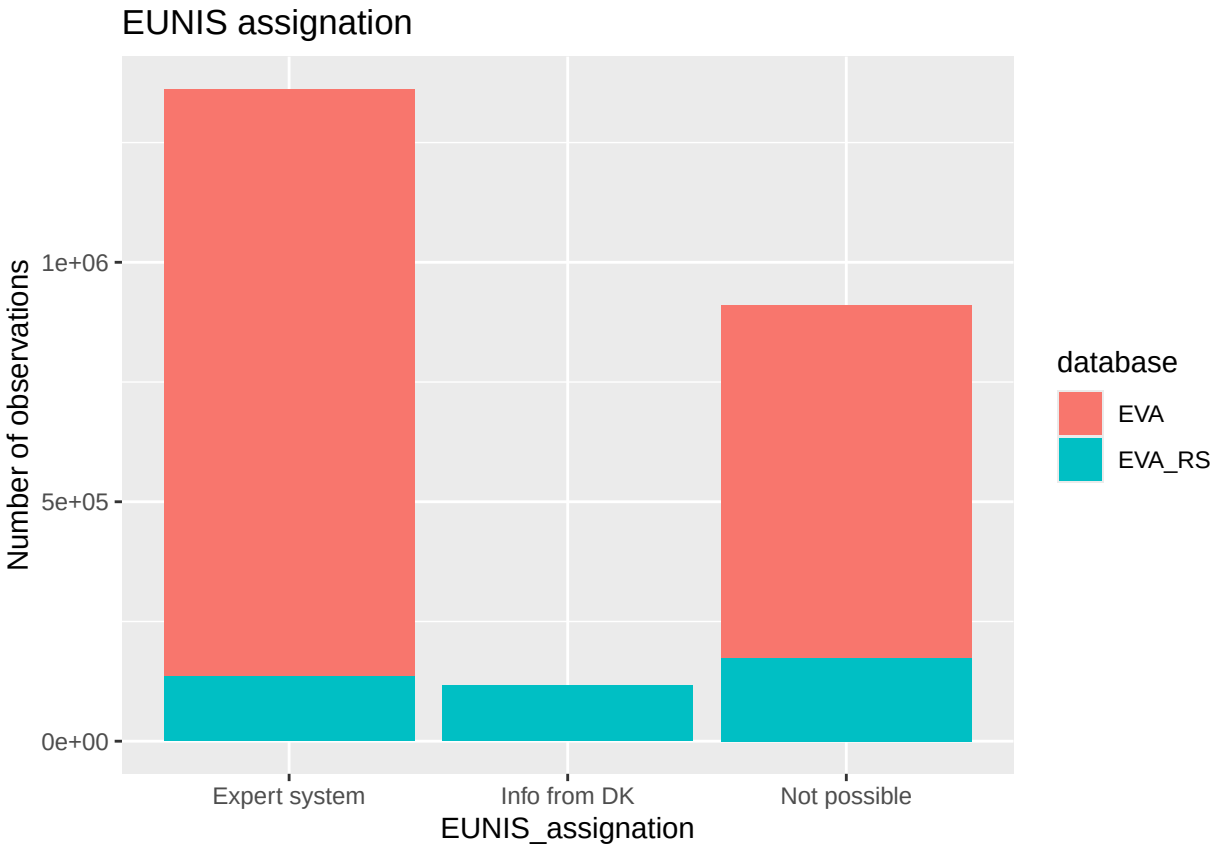
Select key columns and add “database” column

Combine the two databases

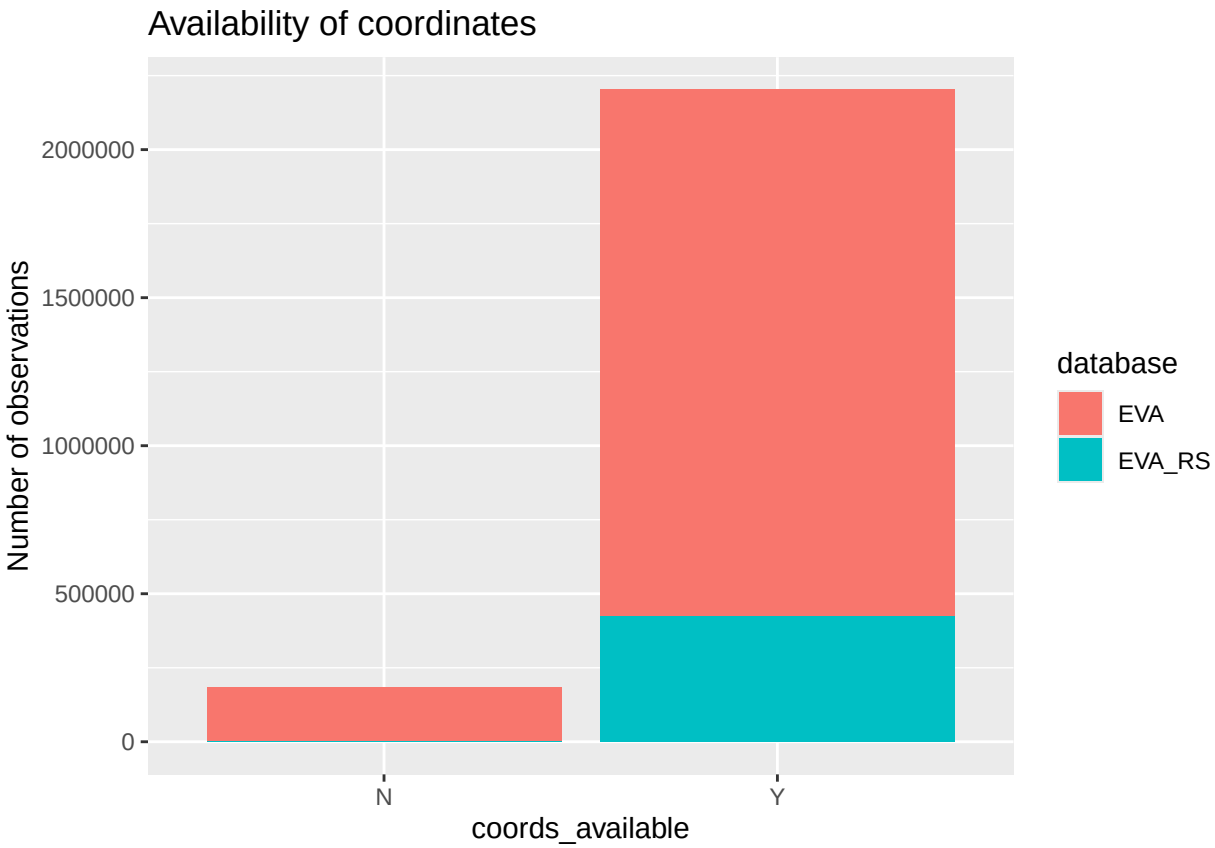
EUNIS level 1 availability



EUNIS assignation

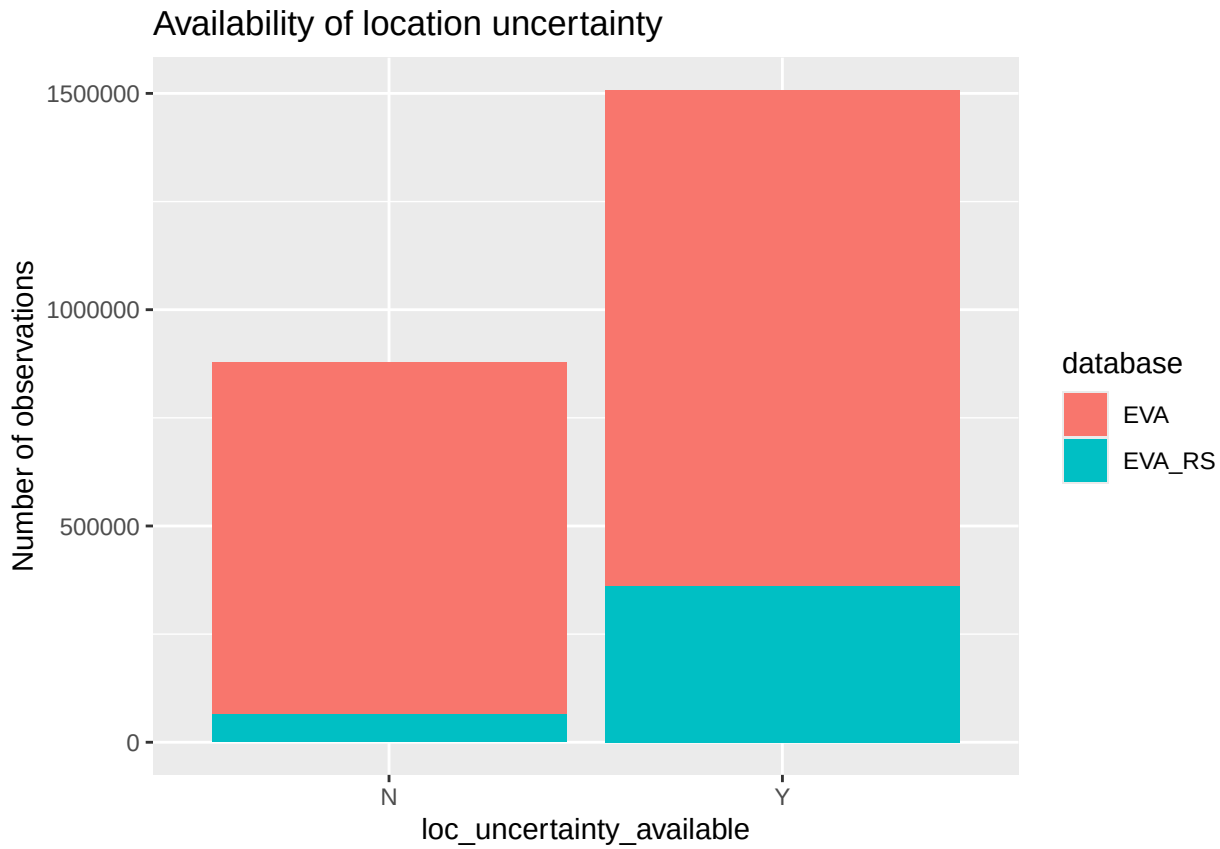


Coordinates available

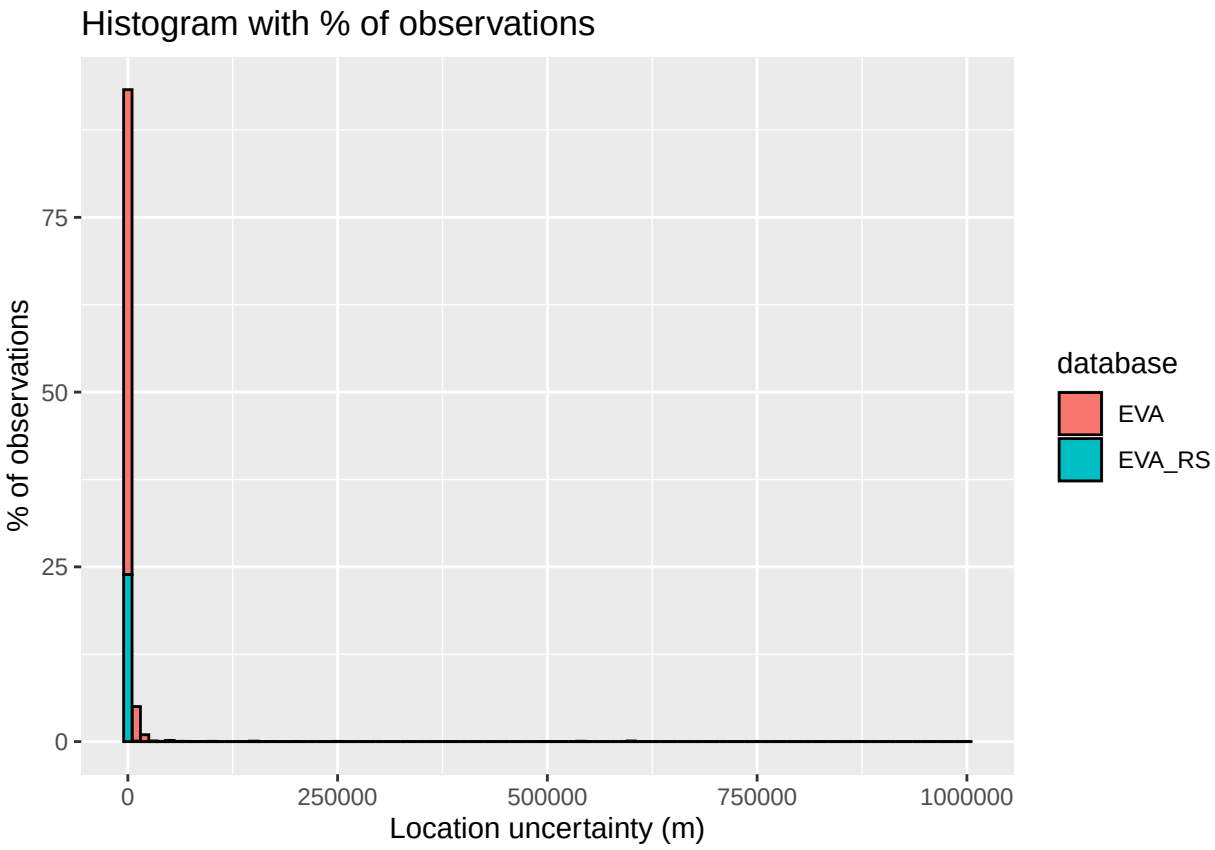


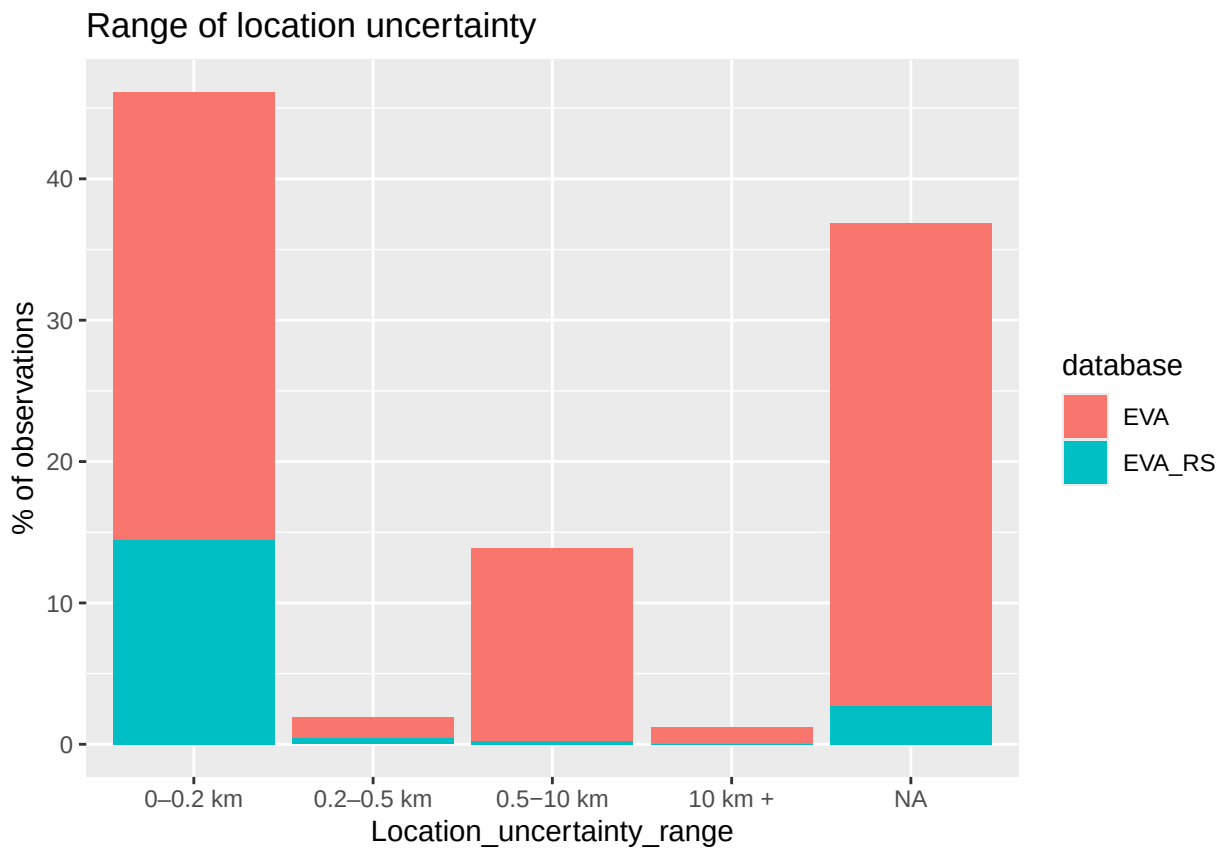
Location uncertainty

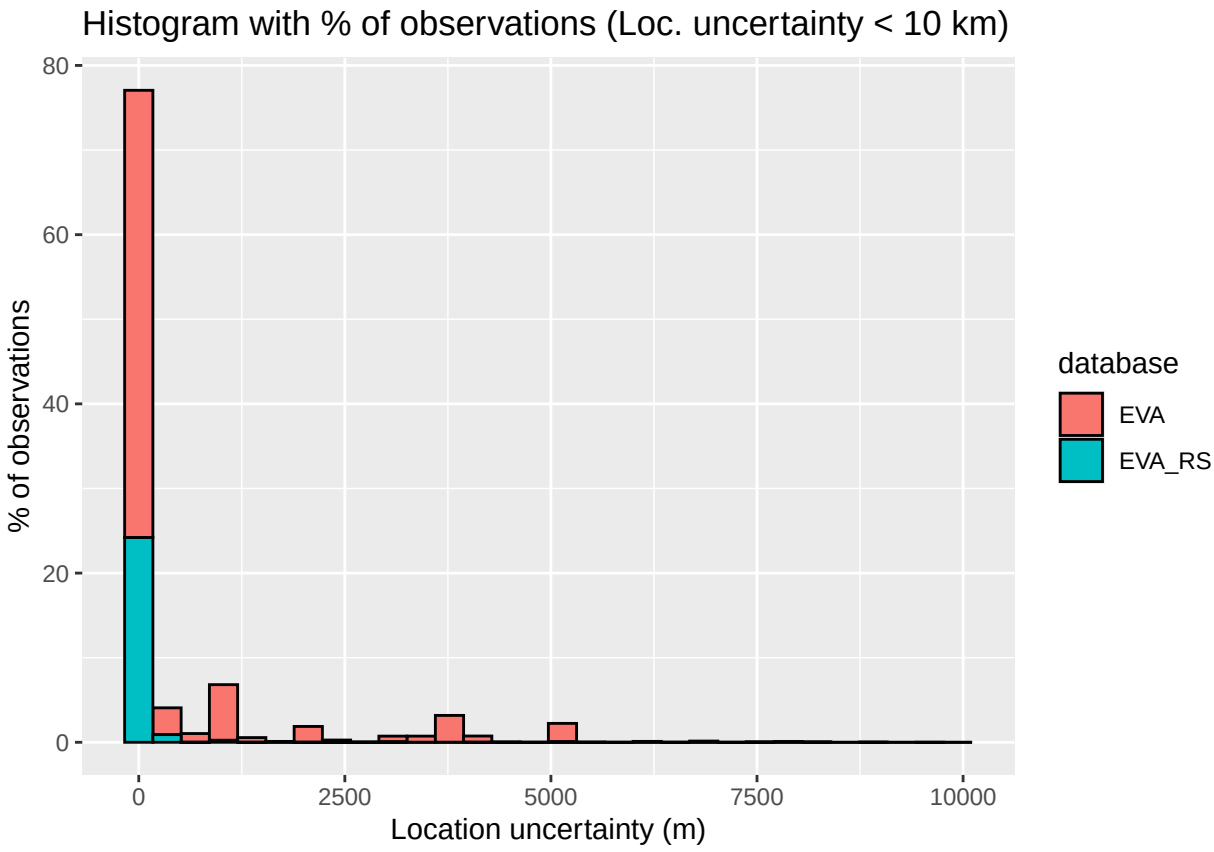
Availabilty

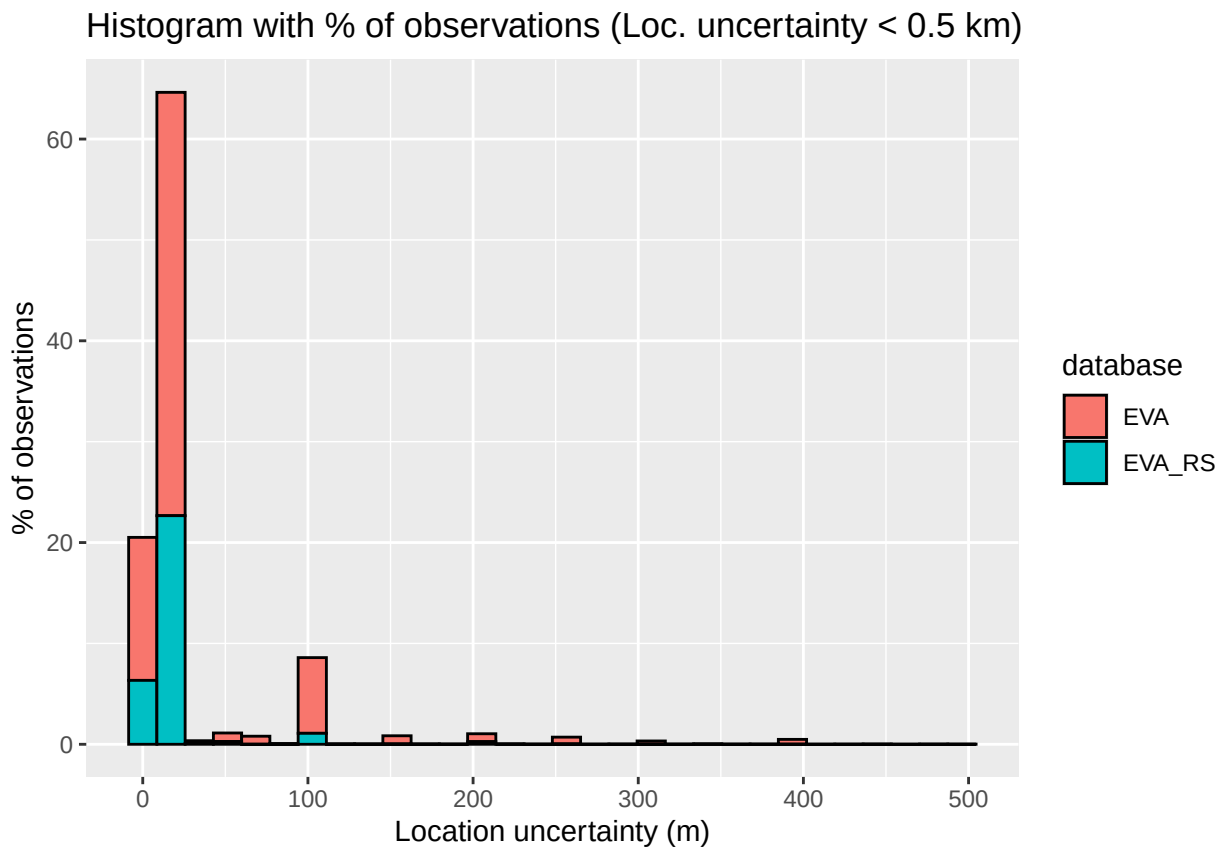


Distribution when available




Range of location uncertainty
% of observations
40
30
20
10
0
0–0.2 km
0.2–0.5 km
0.5–10 km
10 km +
NA
Location_uncertainty_range
database
EVA
EVA_RS


Histogram with % of observations (Loc. uncertainty < 10 km)
% of observations
80
60
40
20
0
0
2500
5000
7500
10000
Location uncertainty (m)
database
EVA
EVA_RS



TO DO: Location uncertainty (last resurveys for EVA_RS)

Extract last resurvey for each each ReSurvey plot (defined by combination of RS_CODE, ReSurvey site and ReSurvey plot). So far, discard 111 resurvey observations corresponding to ReSurvey project "Swiss arable flora since the 1920s" (RS_CODE = CH_0002) where ReSurvey plot is NA.

Careful! There are plots with muliple records for the last year (and also for other years). Solve this and check location uncertainty for last resurveys later on.

Check for plots with multiple records for the same year

Save in file (fewer columns for internal use):

Save in file (all columns to send to Ilona):

Keep GOOD observations WP3

Defining "good" observations as those that have an EUNIS assigned (any level) and where location uncertainty is available and < 0.5 km.

Proportion of "good" observations:

```
## [1] 0.2313703
```

Proportion of observations with EUNIS assigned, irrespective of location uncertainty:

```
## [1] 0.6189554
```

Maybe include those without location uncertainty but where location methods is GPS, etc.?

Check for plots with multiple records for the same year in "good" observations

Keep GOOD observations that do not have multiple records for the same year

Proportion of "good" observations:

```
## [1] 0.2293444
```

Session info

```
## R version 4.4.2 (2024-10-31 ucrt)
## Platform: x86_64-w64-mingw32/x64
## Running under: Windows 10 x64 (build 19045)
##
## Matrix products: default
##
##
## locale:
## [1] LC_COLLATE=Spanish_Spain.utf8  LC_CTYPE=Spanish_Spain.utf8
## [3] LC_MONETARY=Spanish_Spain.utf8 LC_NUMERIC=C
## [5] LC_TIME=Spanish_Spain.utf8
##
## time zone: Europe/Paris
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] readxl_1.4.3    purrr_1.0.2     stringr_1.5.1   ggplot2_3.5.1
## [5] lubridate_1.9.4 dplyr_1.1.4     tidyr_1.3.1     readr_2.1.5
## [9] here_1.0.1
##
## loaded via a namespace (and not attached):
##  [1] bit_4.5.0.1       gtable_0.3.6      crayon_1.5.3      compiler_4.4.2
##  [5] tidyselect_1.2.1  parallel_4.4.2    scales_1.3.0      yaml_2.3.10
##  [9] fastmap_1.2.0     R6_2.5.1          labeling_0.4.3    generics_0.1.3
## [13] knitr_1.49        tibble_3.2.1      munsell_0.5.1     rprojroot_2.0.4
## [17] pillar_1.9.0      tzdb_0.4.0        rlang_1.1.4       utf8_1.2.4
## [21] stringi_1.8.4     xfun_0.49         bit64_4.5.2       timechange_0.3.0
```

```
## [25] cli_3.6.3         withr_3.0.2       magrittr_2.0.3    digest_0.6.37
## [29] grid_4.4.2        vroom_1.6.5       hms_1.1.3         lifecycle_1.0.4
## [33] vctrs_0.6.5       evaluate_1.0.1    glue_1.8.0        farver_2.1.2
## [37] cellranger_1.1.0  fansi_1.0.6       colorspace_2.1-1  rmarkdown_2.29
## [41] tools_4.4.2       pkgconfig_2.0.3   htmltools_0.5.8.1
```