Xie_et_all-full_script

Garland Xie

09/01/2022

Preamble

This document was last updated on January 20 2022.

Garland Xie created this annotated R markdown script. The goal of this script is to maximize transparency and reproducibility associated with the following PhD chapter of my dissertation: **Drivers of invasibility in urban meadow restoration (a field study).**

Session info and loading packages

It is important to note that certain functions are preceded by their package sources to help with understanding and reproducing the code (if necessary). This is denoted as "package::function()."

Load packages

```
library(here)           # for creating relative file-paths
library(googlesheets4)  # for importing googlesheet files
library(dplyr)          # for manipulating data
library(knitr)          # for creating tables
library(kableExtra)     # for customizing tables
library(janitor)        # for making 'R-friendly' column names
library(ggplot2)        # for visualizing data
library(ggrepel)        # for making easy-to-read text labels
library(vegan)          # for analyzing community data matrices
library(patchwork)
```

Session Info

Key information for reproducibility

```
sessionInfo()
```

```
## R version 4.1.1 (2021-08-10)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
```

```
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_CA.UTF-8/en_CA.UTF-8/en_CA.UTF-8/C/en_CA.UTF-8/en_CA.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] patchwork_1.1.1     vegan_2.5-7         lattice_0.20-45
##  [4] permute_0.9-5       ggrepel_0.9.1       ggplot2_3.3.5
##  [7] janitor_2.1.0       kableExtra_1.3.4    knitr_1.34
## [10] dplyr_1.0.7         googlesheets4_1.0.0 here_1.0.1
##
## loaded via a namespace (and not attached):
##  [1] tidyselect_1.1.1  xfun_0.26         purrr_0.3.4       splines_4.1.1
##  [5] gargle_1.2.0      snakecase_0.11.0  colorspace_2.0-2  vctrs_0.3.8
##  [9] generics_0.1.0    htmltools_0.5.2   viridisLite_0.4.0 mgcv_1.8-37
## [13] yaml_2.2.1        utf8_1.2.2        rlang_0.4.11      pillar_1.6.2
## [17] glue_1.4.2        withr_2.4.2       DBI_1.1.1         lifecycle_1.0.0
## [21] stringr_1.4.0     munsell_0.5.0     gtable_0.3.0      cellranger_1.1.0
## [25] rvest_1.0.1       evaluate_0.14     fastmap_1.1.0     parallel_4.1.1
## [29] fansi_0.5.0       Rcpp_1.0.7        scales_1.1.1      webshot_0.5.2
## [33] fs_1.5.0          systemfonts_1.0.2 digest_0.6.28     stringi_1.7.4
## [37] rprojroot_2.0.2   grid_4.1.1        tools_4.1.1       magrittr_2.0.1
## [41] tibble_3.1.4      cluster_2.1.2     crayon_1.4.1      pkgconfig_2.0.3
## [45] Matrix_1.3-4      MASS_7.3-54       ellipsis_0.3.2    xml2_1.3.2
## [49] googledrive_2.0.0 lubridate_1.7.10  assertthat_0.2.1  rmarkdown_2.11
## [53] svglite_2.0.0     httr_1.4.2        rstudioapi_0.13   R6_2.5.1
## [57] nlme_3.1-153      compiler_4.1.1
```

Seed bank

Methods

For the spring sampling season, we collected soil samples for 81 1-m2 plots across nine sites in three different management regimes (i.e., seed drilling, undisturbed, and maintenance mow) in May 2021. Here, we located the centroid of each plot in the field using a handheld GPS. At each plot, we measured a distance of 30 cm away from the centroid. Before sampling, we removed the surface litter layer or gravel. I then took a soil sample using a 3.5-cm diameter soil core to a depth of 5 cm. This procedure was repeated five times to increase the precision and accuracy of the seed bank estimate **(citation)** and reduce the spatial autocorrelation of community abundance and species richness within plots **(citation)**. I combined all five soil samples to one composite seed bank sample per plot into a plastic bag. It is possible to obtain a complete picture of the community dynamics of the seed bank by sampling twice within a growing season such as the spring and fall season **(citation)**. I have also repeated the field measurement protocol for obtaining seed bank samples from October-November 2021. Because fall samples require cold temperatures to break winter dormancy, all spring and fall samples were stored in a freezer at -20 ℃ for up to three weeks prior to processing at the greenhouse.

Prior to seed emergence methods, we processed soils to remove any litter, roots, rocks, or other debris using a 4 mm sieve. We then filled 81 plastic germination trays with 3 cm of sterilized potting soil. I then spread the field-collected soil samples (one composite sample per plot amounting to 240 cubic centimeter of soil)

evenly over the potting soil to ensure complete germination. To serve as a control to track any contaminated seeds from either the greenhouse environment or the potting soil, we filled five additional trays with sterilized potting soil and four additional trays with unsterilized potting (n = 9 for control treatments). All of the trays were randomly dispersed throughout the greenhouse bench. To reduce any spatial effects of environmental heterogeneity (i.e., light intensity and temperature), we completely randomized the trays once a week. We also watered all the trays twice a week to ensure adequate moisture.

We monitored seed germination over a period of four months (July to October 2021). During this time, we scarified the soil if there was at least a week of no seed germination. Any species that were difficult to identify as either a seedling or a juvenile was also collected and placed into a separate tray and grown until flowering. After the fourth month, we removed all the remaining germinants. While previous seed bank study suggests a monitoring period of six months is required, we were primarily interested in species that germinated first and most abundantly as these have competitive advantage in establishment as propagule pressure for both exotic and native plant species. We plan to replicate the seedling emergence for the fall sampling season starting from November 2021 to February 2022.

Import data

Read in data from a Google spreadsheet API

```
# spring sampling season
sb_spr_link <- "https://docs.google.com/spreadsheets/d/1O1Ll_PsW3qKwdZ_xnTrDKT_kGnpLGvtUKBQ73zvvBBM/edi

seed_bank_spr <- googlesheets4::read_sheet(
    sb_spr_link,
    sheet = "raw_data"
    )
```

```
# fall sampling season
sb_fall_link <- "https://docs.google.com/spreadsheets/d/1SWlk5eWdk3IOMFS9nv61p1HW-Tw4K3yR-tAIkAMlzs4/ed

seed_bank_fall <- googlesheets4::read_sheet(
  sb_fall_link,
  sheet = "raw_data"
  )
```

```
# binomial latin names
taxon_link <- "https://docs.google.com/spreadsheets/d/1r-T9lY1Osjez8SKKoCUHRi5ftbHBvgFn2Xk-IPqc0z4/edit

sb_taxa <- googlesheets4::read_sheet(
  taxon_link,
  sheet = "raw_data"
  )
```

Metadata

Descriptions of raw data files:

- **seed_bank_spring** - seedling emergent data from spring season and sampled in the greenhouse from July to October 2021
- **seed_bank_fall** - seedling emergent data from fall season and sampled in the greenhouse from July to October 2021

Descriptions of each variable in the data frames:

- **season**: a factor variable that indicates the sampling season (i.e., Spring or Fall)
- **section:** a factor variable that represents a specific section in the Meadoway (*i.e.*, section 2 or section 4)
- **site:** a factor variable that represents a sampled site (40m X 40m area)
- **treatment:** a factor variable that represents a management regime (*i.e.*, tilling, restored, or mowing) or control treatments (*i.e.*, potting soil, sterilized potting soil)
- **plot:** a factor variable represent a sampled plot within a site
- **date_ymd:** a date of when seedling emergent was identified and recorded (using ISO 8601 format)
- **spp_code**: a character string that represents an abbreviation of the taxonomic species name **(see Table X)**.
- **abund:** an integer variable that represents the seedling emergent abundance for a specific plot
- **sampled_by**: a character string that represents the initials of the surveyor
- **comments**: additional comments

Data wrangling

Reorganize data. For each sampling season, subset the seed bank dataset to six sites that are congruent with the above-ground biomass dataset. Then, merge the subset with the taxonomic data to obtain binomial latin species name (e.g., *Alliara_petiolata).* Finally, combine both fall and spring seed bank data into a single data-frame.

```
sb_spring_tidy <- seed_bank_spr %>%
  janitor::clean_names() %>%
  dplyr::filter(
    site %in% c("BNSH", "GRNB", "DAVE", "KENN", "TIMH", "VICP")
    ) %>%
  left_join(sb_taxa %>% janitor::clean_names(), by = c("spp_code" = "code")) %>%
  mutate(binom_latin = paste(genus, species, sep = "_")) %>%
  select(
    season,
    section,
    site,
    treatment,
    plot,
    date_ymd,
```

```
    spp_code,
    binom_latin,
    abund
  )

sb_fall_tidy <- seed_bank_fall %>%
  janitor::clean_names() %>%
  dplyr::filter(
    site %in% c("BNSH", "GRNB", "DAVE", "KENN", "TIMH", "VICP")
    ) %>%
  left_join(sb_taxa %>% janitor::clean_names(), by = c("spp_code" = "code")) %>%
  mutate(binom_latin = paste(genus, species, sep = "_")) %>%
  select(
    season,
    section,
    site,
    treatment,
    plot,
    date_ymd,
    spp_code,
    binom_latin,
    abund
  )

sb_tidy <- rbind(sb_fall_tidy, sb_spring_tidy)
```

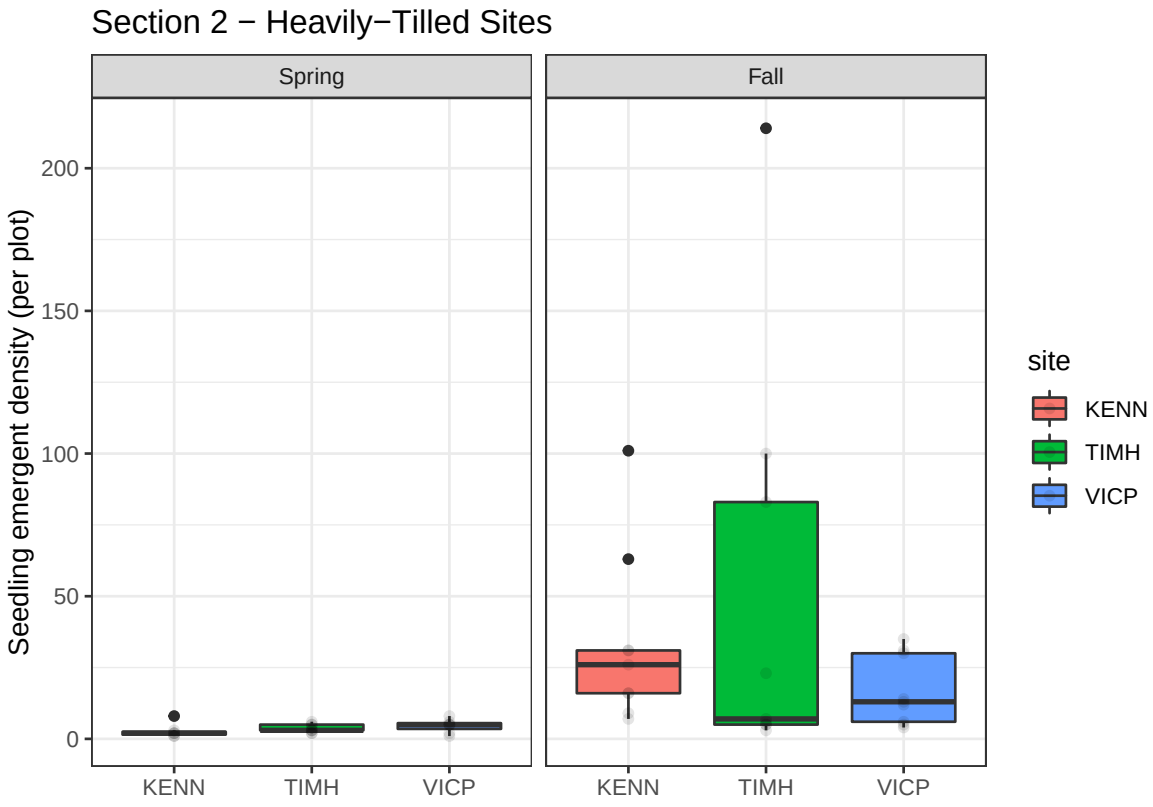
Summarize data. The raw data for seedling emergent abundance was collected at specific dates (y-m-d) within specific plots. Temporal dimensions are beyond the scope of this analysis, and the data will thus be summarized as **total seedling emergent density per plot**. This measured variable make sense since we collected soil cores within a defined volume (3.5 cm diameter and a 5 cm depth, approximating the shape of a cylinder).

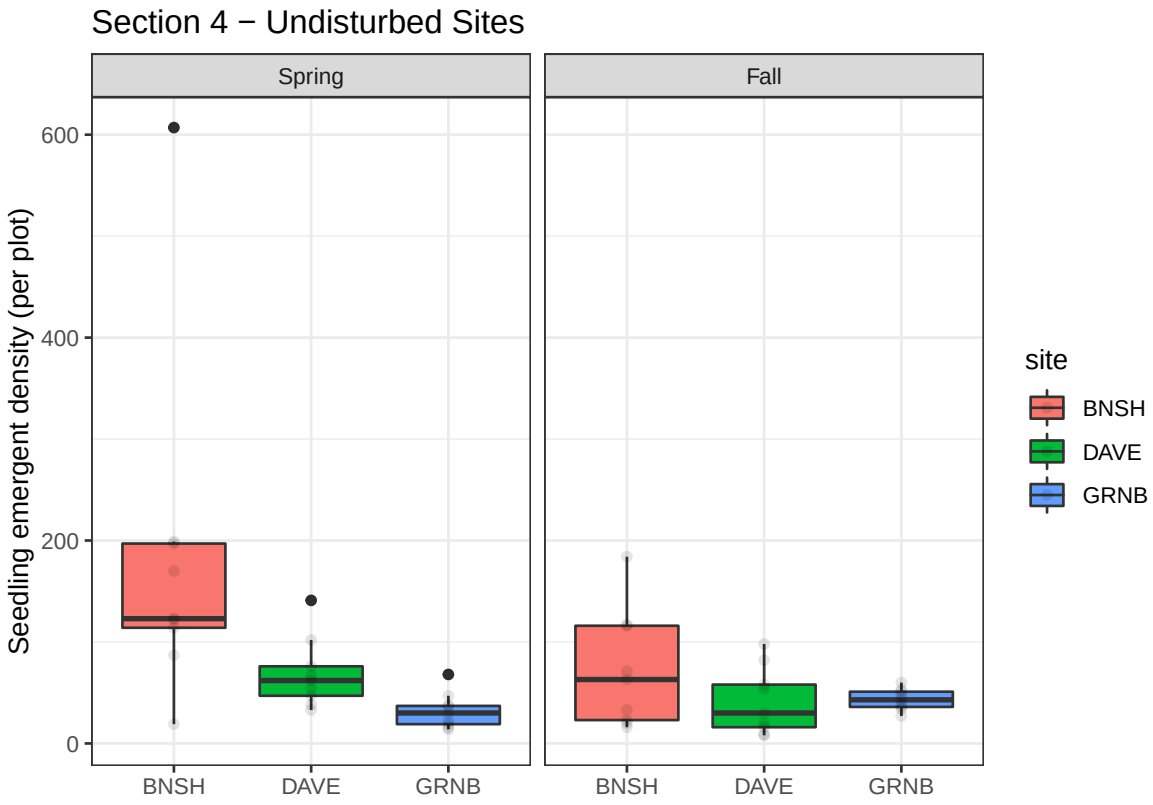
```
sb_summary <- sb_tidy %>%
  group_by(season, section, site, treatment, plot) %>%
  summarize(
    seed_emer_density = sum(abund, na.rm = TRUE)
  ) %>%
  ungroup()
```

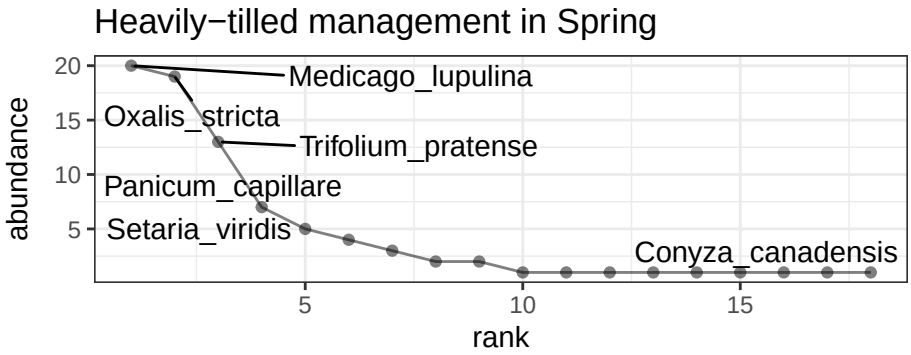
Exploratory data analysis

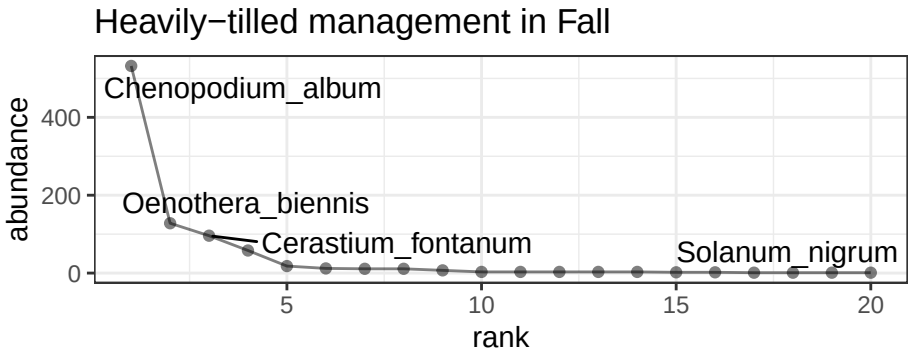
Seed Bank Analysis

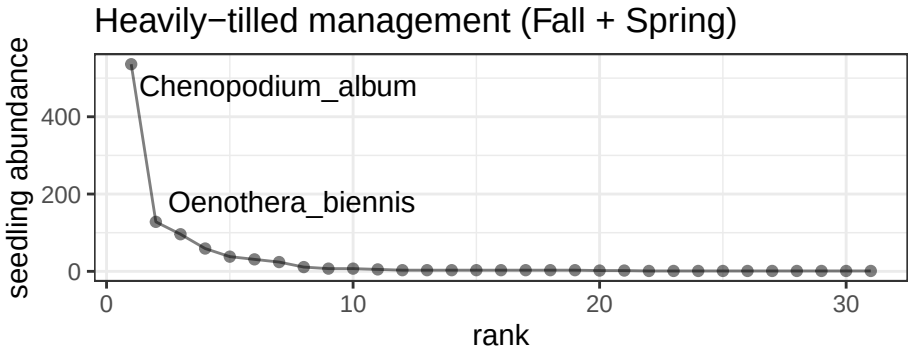
- How many seeds were in the seed bank across both sampling season?4935 seeds.
- How many seeds were in the seed bank for the fall season? 2303 seeds.
- How many seeds were in the seed bank in the spring season?2632 seeds.
- How does seed bank density differ across sites? See figure below. Note that each point represent a particular plot.

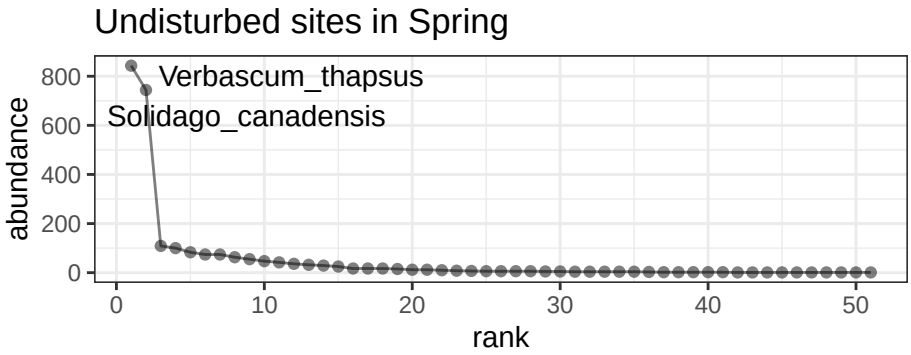

Section 2 − Heavily−Tilled Sites
Spring
Fall
Seedling emergent density (per plot)
0
50
100
150
200
KENN
TIMH
VICP
site
KENN
TIMH
VICP

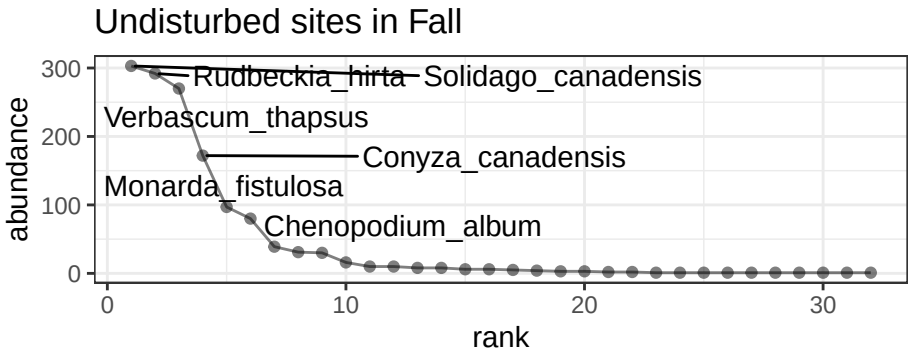

Section 4 − Undisturbed Sites
Spring
Fall
Seedling emergent density (per plot)
0
200
400
600
BNSH
DAVE
GRNB
site
BNSH
DAVE
GRNB

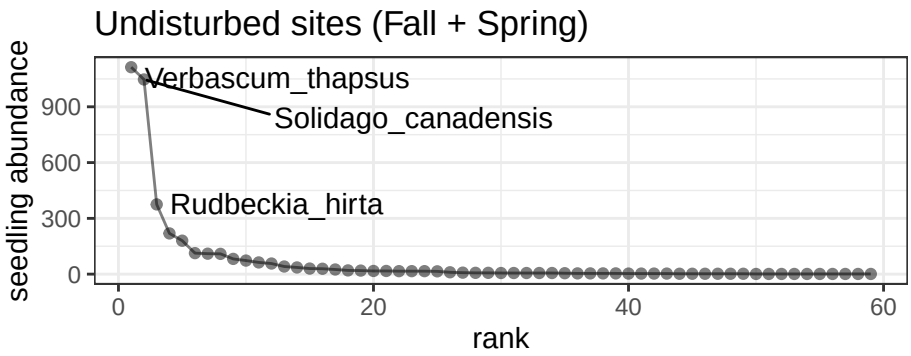

Heavily−tilled management in Spring
abundance
rank
Medicago_lupulina
Oxalis_stricta
Trifolium_pratense
Panicum_capillare
Setaria_viridis
Conyza_canadensis


Heavily−tilled management in Fall
abundance
rank
Chenopodium_album
Oenothera_biennis
Cerastium_fontanum
Solanum_nigrum


Heavily−tilled management (Fall + Spring)
seedling abundance
rank
Chenopodium_album
Oenothera_biennis

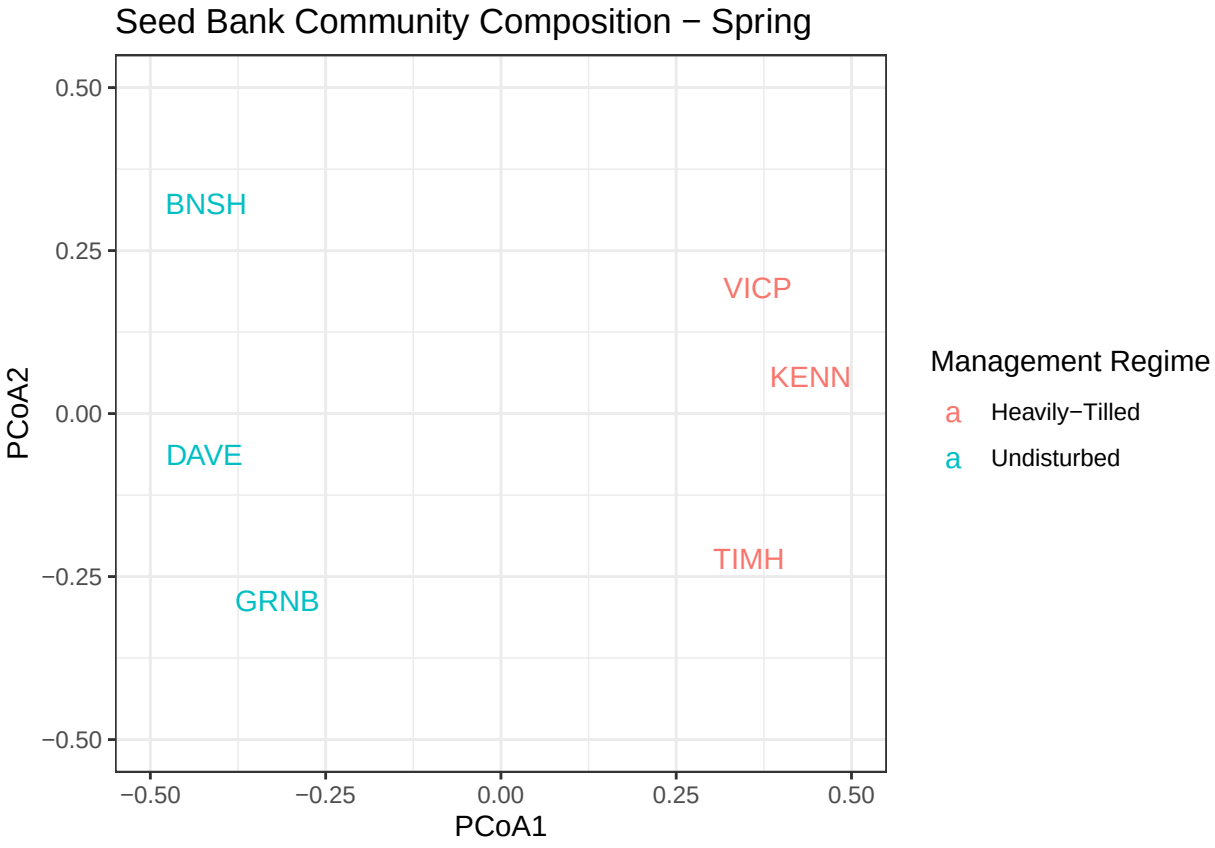

Undisturbed sites in Spring
Verbascum_thapsus
Solidago_canadensis
abundance
rank

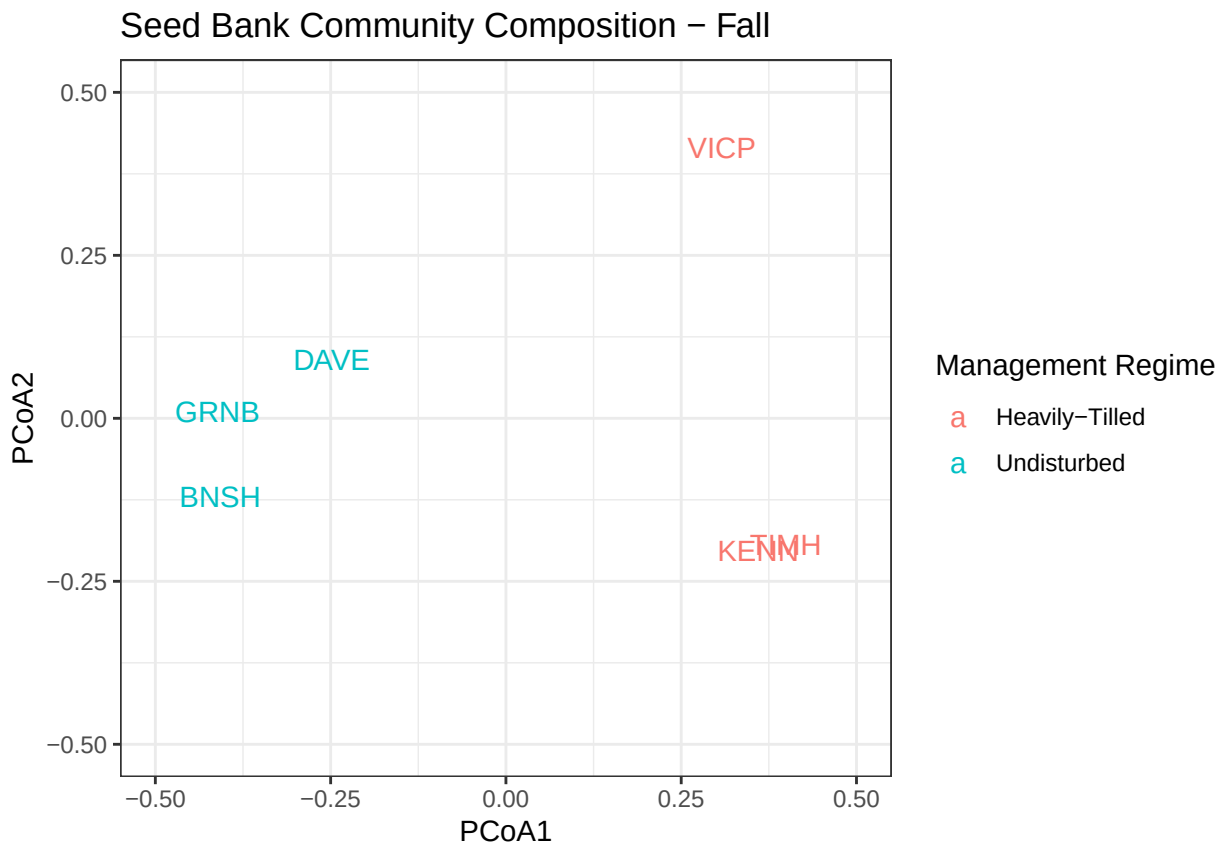

Undisturbed sites in Fall
Rudbeckia_hirta
Solidago_canadensis
Verbascum_thapsus
Conyza_canadensis
Monarda_fistulosa
Chenopodium_album
abundance
rank


Undisturbed sites (Fall + Spring)
Verbascum_thapsus
Solidago_canadensis
Rudbeckia_hirta
seedling abundance
rank

Community composition It would be interesting to explore how the community composition of the seed bank differ among plots, sites, seasons, and management regimes. Here, two possible approaches to answer this question are: (1) principal coordinate analysis (PCoA), and (2) non-metric dimensional scaling (NMDS).

Using a dissimilarity measure (e.g., Bray-Curtis), both can help visualize close objects with similar values (*i.e.,* seed banks with similar community composition are close together) and distance objects with different values. Here, I will follow (Borcard, Gillet, and Legendre 2011) recommendation on using PCoA, instead than NMDS, for this analysis.





References

Borcard, Daniel, François Gillet, and Pierre Legendre. 2011. *Numerical Ecology with R*. Use R! New York: Springer-Verlag. https://doi.org/10.1007/978-1-4419-7976-6.