

Format covariates

Aidan Brushett

2025-05-02

Contents

Before you begin	2
0. Setup	2
1. Import all covariate data	2
2. Tidy up the HFI data	3
2.1. Initial cleaning, add the sample year column	3
1.2. Group the HFI data	10
1.3. Group the harvest data by year	13
1.4. Add the harvest and do some additional grouping	14
3. Tidy up the forest inventory data	15
3.1. Clean up and add sample year column	15
3.2. Calculate fire by year columns	16
3.3. Finalize and give the species a more meaningful name	17
4. Format the configuration data	17
4.1. Merge data and select the metrics we want	17
5. Put it all together:	19
5.1. Join the data	19
5.2. Append lat/long and UTM 12N coordinates	19
5.3. Final inspection	19
5.4. Save the data:	22

Before you begin

This script is number 2 of 6 in a series of scripts used to replicate the analyses presented in the paper: “Life on the edge: Industrial footprint and edge effects variably affect the distribution of a boreal small mammal”

This script was used to append, combine, and rename the landscape metrics for all sites in the study that were extracted in the previous script. **The raw spatial data used to develop landcover metrics is not available on GitHub but can be shared by the authors upon request.**

When running these scripts, please ensure that you have downloaded the complete GitHub repository. This will ensure you have all the files, data, and proper folder structure you will need to run this code and associated analyses.

Also make sure you open RStudio through the R project (OSM_red_squirrel_distribution.Rproj). This will automatically set your working directory to the correct place (wherever you saved the repository) and ensure you don't have to change the file paths for some of the data. This analysis was initially run in R v4.3.0. If you have any questions or concerns, please contact one of the authors (in order):

Aidan Brushett M.Sc. Student University of Victoria
School of Environmental Studies
Email: aidanbrushett@uvic.ca

Emerald Arthurs M.Sc. Student University of Victoria
School of Environmental Studies

0. Setup

```
library(sf)
library(tidyverse)
rm(list=ls())
```

1. Import all covariate data

```
hfi <- read_csv("./data/raw/OSM_HFI2021_metrics.csv")
```

```
## Rows: 9460 Columns: 134
## -- Column specification -----
## Delimiter: ","
## chr  (2): array, site
## dbl (132): buffer_dist, AIRP-RUNWAY, BORROWPIT-DRY, BORROWPIT-WET, BORROWPIT...
##
## 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.
```

```
sbfi <- read_csv("./data/raw/OSM_SBF2020_metrics.csv")
```

```
## Rows: 9460 Columns: 109
## -- Column specification -----
## Delimiter: ","
## chr (2): array, site
## dbl (107): buffer_dist, AGE_0_10, AGE_10_20, AGE_20_30, AGE_30_40, AGE_40_50...
##
## 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.
```

```
config1 <- read_csv("../data/raw/OSM_simple_config_landscapemetrics.csv")
```

```
## Rows: 9460 Columns: 15
## -- Column specification -----
## Delimiter: ","
## chr (1): site
## dbl (11): buffer, water_cai_mn, natural_cai_mn, anthropogenic_cai_mn, water...
## lgl (3): landscape_cai_mn, landscape_ed, landscape_tca
##
## 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.
```

```
config2 <- read_csv("../data/raw/OSM_grouped_config_landscapemetrics.csv")
```

```
## Rows: 9460 Columns: 8
## -- Column specification -----
## Delimiter: ","
## chr (1): site
## dbl (7): buffer, landscape_cohesion, landscape_contag, landscape_mesh, lands...
##
## 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.
```

2. Tidy up the HFI data

2.1. Initial cleaning, add the sample year column

```
hfi <- hfi %>%

  set_names(tolower(names(.)) %>%
    str_replace_all(., pattern = c('-'), # provide the character pattern to look for (if
    ↪ you don't keep the \\ it won't work)
    replacement = '_') %>% # what you want the pattern to be replaced
    ↪ with
    str_replace_all(., pattern = c('\\.'), # provide the character pattern to look for
    ↪ (if you don't keep the \\ it won't work)
    replacement = '_')
  ) %>%

  # What year were the arrays sampled? Used to determine recent fire and burns.
  # Currently, entered manually since this is the most flexible option.
```

```
mutate(
  array_year = case_when(
    array %in% c("LU2", "LU3") ~ 2021,
    array %in% c("LU13", "LU15", "LU21", "LU1") ~ 2022,
    array %in% c("LU9", "LU16", "LU14", "LU22") ~ 2023)
)

summary(hfi)
```

```
##      array      site      buffer_dist      airp_runway
## Length:9460      Length:9460      Min.   : 50      Min.   :0.000e+00
## Class :character      Class :character      1st Qu.:1000      1st Qu.:0.000e+00
## Mode  :character      Mode  :character      Median :2375      Median :0.000e+00
##                                         Mean  :2393      Mean  :3.799e-05
##                                         3rd Qu.:3750      3rd Qu.:0.000e+00
##                                         Max.   :5000      Max.   :2.322e-02
## borrowpit_dry      borrowpit_wet      borrowpits
## Min.   :0.0000000      Min.   :0.0000000      Min.   :0.0000000
## 1st Qu.:0.0000000      1st Qu.:0.0000000      1st Qu.:0.0000000
## Median :0.0000000      Median :0.0000000      Median :0.0000000
## Mean   :0.0009587      Mean   :0.0009804      Mean   :0.0005969
## 3rd Qu.:0.0004827      3rd Qu.:0.0002791      3rd Qu.:0.0000336
## Max.   :0.2655422      Max.   :0.5356491      Max.   :1.0000000
## camp_industrial      campground      canal
## Min.   :0.0000000      Min.   :0.000e+00      Min.   :0.000e+00
## 1st Qu.:0.0000000      1st Qu.:0.000e+00      1st Qu.:0.000e+00
## Median :0.0000000      Median :0.000e+00      Median :0.000e+00
## Mean   :0.0005183      Mean   :5.073e-05      Mean   :8.226e-06
## 3rd Qu.:0.0000000      3rd Qu.:0.000e+00      3rd Qu.:0.000e+00
## Max.   :0.2450556      Max.   :2.897e-02      Max.   :1.961e-02
## cfo      clearing_unknown      clearing_wellpad_unconfirmed
## Min.   :0.000e+00      Min.   :0.0000000      Min.   :0.0000000
## 1st Qu.:0.000e+00      1st Qu.:0.0000000      1st Qu.:0.0000000
## Median :0.000e+00      Median :0.000132      Median :0.0000000
## Mean   :2.659e-07      Mean   :0.003893      Mean   :0.0005047
## 3rd Qu.:0.000e+00      3rd Qu.:0.002124      3rd Qu.:0.0002556
## Max.   :1.217e-03      Max.   :0.507481      Max.   :0.4761982
## conventional_seismic      country_residence      crop
## Min.   :0.000000      Min.   :0.000e+00      Min.   :0.000e+00
## 1st Qu.:0.002977      1st Qu.:0.000e+00      1st Qu.:0.000e+00
## Median :0.006318      Median :0.000e+00      Median :0.000e+00
## Mean   :0.008086      Mean   :7.716e-05      Mean   :7.260e-07
## 3rd Qu.:0.009888      3rd Qu.:0.000e+00      3rd Qu.:0.000e+00
## Max.   :0.151463      Max.   :2.505e-02      Max.   :2.571e-03
## cultivation_abandoned      dugout      facility_other
## Min.   :0.000e+00      Min.   :0.000e+00      Min.   :0.0000000
## 1st Qu.:0.000e+00      1st Qu.:0.000e+00      1st Qu.:0.0000000
## Median :0.000e+00      Median :0.000e+00      Median :0.0000000
## Mean   :1.255e-05      Mean   :1.714e-06      Mean   :0.0003826
## 3rd Qu.:0.000e+00      3rd Qu.:0.000e+00      3rd Qu.:0.0000000
## Max.   :3.115e-02      Max.   :1.825e-03      Max.   :0.2009921
## facility_unknown      greenspace      grvl_sand_pit      harvest_area_1940
## Min.   :0.0000000      Min.   :0.000e+00      Min.   :0.0000000      Min.   :0.000e+00
## 1st Qu.:0.0000000      1st Qu.:0.000e+00      1st Qu.:0.0000000      1st Qu.:0.000e+00
```

##	Median :0.0000000	Median :0.000e+00	Median :0.0000000	Median :0.000e+00
##	Mean :0.0002191	Mean :8.281e-06	Mean :0.001667	Mean :9.983e-06
##	3rd Qu.:0.0000000	3rd Qu.:0.000e+00	3rd Qu.:0.0000000	3rd Qu.:0.000e+00
##	Max. :0.1379450	Max. :3.119e-03	Max. :0.557859	Max. :5.278e-03
##	harvest_area_1950	harvest_area_1960	harvest_area_1963	
##	Min. :0.000e+00	Min. :0.0000000	Min. :0.000e+00	
##	1st Qu.:0.000e+00	1st Qu.:0.0000000	1st Qu.:0.000e+00	
##	Median :0.000e+00	Median :0.0000000	Median :0.000e+00	
##	Mean :1.023e-05	Mean :0.0005155	Mean :2.994e-05	
##	3rd Qu.:0.000e+00	3rd Qu.:0.0000000	3rd Qu.:0.000e+00	
##	Max. :8.774e-03	Max. :0.1025133	Max. :4.926e-02	
##	harvest_area_1970	harvest_area_1971	harvest_area_1972	harvest_area_1973
##	Min. :0.0000000	Min. :0.0000000	Min. :0.0000000	Min. :0.0000000
##	1st Qu.:0.0000000	1st Qu.:0.0000000	1st Qu.:0.0000000	1st Qu.:0.0000000
##	Median :0.0000000	Median :0.0000000	Median :0.0000000	Median :0.0000000
##	Mean :0.0005073	Mean :0.0004166	Mean :0.000198	Mean :0.0000292
##	3rd Qu.:0.0000000	3rd Qu.:0.0000000	3rd Qu.:0.0000000	3rd Qu.:0.0000000
##	Max. :0.1199953	Max. :0.6428582	Max. :0.072798	Max. :0.0123655
##	harvest_area_1975	harvest_area_1976	harvest_area_1977	harvest_area_1978
##	Min. :0.0000000	Min. :0.0000000	Min. :0.0000000	Min. :0.0000000
##	1st Qu.:0.0000000	1st Qu.:0.0000000	1st Qu.:0.0000000	1st Qu.:0.0000000
##	Median :0.0000000	Median :0.0000000	Median :0.0000000	Median :0.0000000
##	Mean :0.0002961	Mean :0.0002204	Mean :0.0008288	Mean :0.000971
##	3rd Qu.:0.0000000	3rd Qu.:0.0000000	3rd Qu.:0.0000000	3rd Qu.:0.0000000
##	Max. :0.1129171	Max. :0.0692077	Max. :1.0000000	Max. :0.930748
##	harvest_area_1979	harvest_area_1980	harvest_area_1984	harvest_area_1985
##	Min. :0.0000000	Min. :0.0000000	Min. :0.0000000	Min. :0.0000000
##	1st Qu.:0.0000000	1st Qu.:0.0000000	1st Qu.:0.0000000	1st Qu.:0.0000000
##	Median :0.0000000	Median :0.0000000	Median :0.0000000	Median :0.0000000
##	Mean :0.0001715	Mean :0.001089	Mean :0.0005546	Mean :0.0003191
##	3rd Qu.:0.0000000	3rd Qu.:0.0000000	3rd Qu.:0.0000000	3rd Qu.:0.0000000
##	Max. :0.0691178	Max. :0.927279	Max. :0.9558170	Max. :0.2627665
##	harvest_area_1986	harvest_area_1987	harvest_area_1988	
##	Min. :0.0000000	Min. :0.0000000	Min. :0.0000000	
##	1st Qu.:0.0000000	1st Qu.:0.0000000	1st Qu.:0.0000000	
##	Median :0.0000000	Median :0.0000000	Median :0.0000000	
##	Mean :0.0001293	Mean :0.0006555	Mean :0.0004326	
##	3rd Qu.:0.0000000	3rd Qu.:0.0000000	3rd Qu.:0.0000000	
##	Max. :0.0394281	Max. :0.1979973	Max. :0.0981403	
##	harvest_area_1989	harvest_area_1990	harvest_area_1991	harvest_area_1992
##	Min. :0.0000000	Min. :0.000e+00	Min. :0.0000000	Min. :0.0000000
##	1st Qu.:0.0000000	1st Qu.:0.000e+00	1st Qu.:0.0000000	1st Qu.:0.0000000
##	Median :0.0000000	Median :0.000e+00	Median :0.0000000	Median :0.0000000
##	Mean :0.0001448	Mean :4.777e-05	Mean :0.001799	Mean :0.0005736
##	3rd Qu.:0.0000000	3rd Qu.:0.000e+00	3rd Qu.:0.0000000	3rd Qu.:0.0000000
##	Max. :0.0330899	Max. :2.986e-02	Max. :1.0000000	Max. :0.1706980
##	harvest_area_1993	harvest_area_1994	harvest_area_1995	harvest_area_1996
##	Min. :0.0000000	Min. :0.0000000	Min. :0.0000000	Min. :0.000e+00
##	1st Qu.:0.0000000	1st Qu.:0.0000000	1st Qu.:0.0000000	1st Qu.:0.000e+00
##	Median :0.0000000	Median :0.0000000	Median :0.0000000	Median :0.000e+00
##	Mean :0.003706	Mean :0.001072	Mean :0.001488	Mean :2.235e-05
##	3rd Qu.:0.0000000	3rd Qu.:0.0000000	3rd Qu.:0.0000000	3rd Qu.:0.000e+00
##	Max. :1.0000000	Max. :0.420749	Max. :0.193473	Max. :1.192e-02
##	harvest_area_1997	harvest_area_1998	harvest_area_1999	harvest_area_2000

## Min. :0.000000	Min. :0.000000	Min. :0.00000	Min. :0.0000000
## 1st Qu.:0.000000	1st Qu.:0.000000	1st Qu.:0.00000	1st Qu.:0.0000000
## Median :0.000000	Median :0.000000	Median :0.00000	Median :0.0000000
## Mean :0.001618	Mean :0.001201	Mean :0.00017	Mean :0.0005354
## 3rd Qu.:0.000000	3rd Qu.:0.000000	3rd Qu.:0.00000	3rd Qu.:0.0000000
## Max. :0.927641	Max. :0.858142	Max. :0.03352	Max. :0.0955988
## harvest_area_2001	harvest_area_2002	harvest_area_2003	harvest_area_2004
## Min. :0.000000	Min. :0.0000000	Min. :0.000000	Min. :0.000000
## 1st Qu.:0.000000	1st Qu.:0.0000000	1st Qu.:0.000000	1st Qu.:0.000000
## Median :0.000000	Median :0.0000000	Median :0.000000	Median :0.000000
## Mean :0.001017	Mean :0.0003775	Mean :0.003453	Mean :0.007816
## 3rd Qu.:0.000000	3rd Qu.:0.0000000	3rd Qu.:0.000000	3rd Qu.:0.000000
## Max. :1.000000	Max. :0.0959808	Max. :0.999935	Max. :0.759461
## harvest_area_2005	harvest_area_2006	harvest_area_2007	harvest_area_2008
## Min. :0.000000	Min. :0.000000	Min. :0.0000000	Min. :0.000000
## 1st Qu.:0.000000	1st Qu.:0.000000	1st Qu.:0.0000000	1st Qu.:0.000000
## Median :0.000000	Median :0.000000	Median :0.0000000	Median :0.000000
## Mean :0.001929	Mean :0.001695	Mean :0.0002326	Mean :0.000719
## 3rd Qu.:0.000000	3rd Qu.:0.000000	3rd Qu.:0.0000000	3rd Qu.:0.000000
## Max. :0.584229	Max. :0.860401	Max. :0.0499629	Max. :0.118363
## harvest_area_2009	harvest_area_2010	harvest_area_2011	harvest_area_2012
## Min. :0.0000000	Min. :0.000000	Min. :0.0000000	Min. :0.000000
## 1st Qu.:0.0000000	1st Qu.:0.000000	1st Qu.:0.0000000	1st Qu.:0.000000
## Median :0.0000000	Median :0.000000	Median :0.0000000	Median :0.000000
## Mean :0.0001108	Mean :0.001124	Mean :0.0004615	Mean :0.002709
## 3rd Qu.:0.0000000	3rd Qu.:0.000000	3rd Qu.:0.0000000	3rd Qu.:0.000000
## Max. :0.0937054	Max. :0.670866	Max. :0.1574343	Max. :0.436828
## harvest_area_2013	harvest_area_2014	harvest_area_2015	harvest_area_2016
## Min. :0.000000	Min. :0.000000	Min. :0.0000000	Min. :0.000000
## 1st Qu.:0.000000	1st Qu.:0.000000	1st Qu.:0.0000000	1st Qu.:0.000000
## Median :0.000000	Median :0.000000	Median :0.0000000	Median :0.000000
## Mean :0.001363	Mean :0.002497	Mean :0.0008442	Mean :0.004194
## 3rd Qu.:0.000000	3rd Qu.:0.000000	3rd Qu.:0.0000000	3rd Qu.:0.000000
## Max. :1.000000	Max. :1.000000	Max. :0.6170907	Max. :1.000000
## harvest_area_2017	harvest_area_2018	harvest_area_2019	harvest_area_2020
## Min. :0.0000000	Min. :0.000e+00	Min. :0.0000000	Min. :0.000000
## 1st Qu.:0.0000000	1st Qu.:0.000e+00	1st Qu.:0.0000000	1st Qu.:0.000000
## Median :0.0000000	Median :0.000e+00	Median :0.0000000	Median :0.000000
## Mean :0.0002134	Mean :8.578e-05	Mean :0.0003999	Mean :0.002193
## 3rd Qu.:0.0000000	3rd Qu.:0.000e+00	3rd Qu.:0.0000000	3rd Qu.:0.000000
## Max. :0.4053079	Max. :4.441e-02	Max. :0.2561170	Max. :0.582707
## harvest_area_2021	harvest_area_white_zone	lagoon	
## Min. :0.0000000	Min. :0.0000000	Min. :0.0000000	
## 1st Qu.:0.0000000	1st Qu.:0.0000000	1st Qu.:0.0000000	
## Median :0.0000000	Median :0.0000000	Median :0.0000000	
## Mean :0.0009868	Mean :0.0001176	Mean :0.0001348	
## 3rd Qu.:0.0000000	3rd Qu.:0.0000000	3rd Qu.:0.0000000	
## Max. :0.3516605	Max. :0.0543438	Max. :0.0261997	
## low_impact_seismic	mines_oilsands	misc_oil_gas_facility	
## Min. :0.0000000	Min. :0.000000	Min. :0.000000	
## 1st Qu.:0.0000000	1st Qu.:0.000000	1st Qu.:0.000000	
## Median :0.0000000	Median :0.000000	Median :0.000000	
## Mean :0.0047371	Mean :0.001381	Mean :0.001836	
## 3rd Qu.:0.0003958	3rd Qu.:0.000000	3rd Qu.:0.000000	

## Max. :0.1144061	Max. :0.398656	Max. :0.389973	
## oil_gas_plant	open_pit_mine	pipeline	reservoir
## Min. :0.000000	Min. :0.000000	Min. :0.000000	Min. :0.000e+00
## 1st Qu.:0.000000	1st Qu.:0.000000	1st Qu.:0.000000	1st Qu.:0.000e+00
## Median :0.000000	Median :0.000000	Median :0.006875	Median :0.000e+00
## Mean :0.001115	Mean :0.000792	Mean :0.015044	Mean :4.108e-06
## 3rd Qu.:0.000000	3rd Qu.:0.000000	3rd Qu.:0.019897	3rd Qu.:0.000e+00
## Max. :0.289878	Max. :0.359047	Max. :0.358825	Max. :7.894e-03
## residence_clearing	ris_borrowpits	ris_camp_industrial	
## Min. :0.000e+00	Min. :0.000e+00	Min. :0.000e+00	
## 1st Qu.:0.000e+00	1st Qu.:0.000e+00	1st Qu.:0.000e+00	
## Median :0.000e+00	Median :0.000e+00	Median :0.000e+00	
## Mean :3.959e-06	Mean :1.432e-05	Mean :2.591e-06	
## 3rd Qu.:0.000e+00	3rd Qu.:0.000e+00	3rd Qu.:0.000e+00	
## Max. :3.113e-03	Max. :5.063e-03	Max. :1.945e-03	
## ris_clearing_unknown	ris_drainage	ris_facility_operations	
## Min. :0.000000	Min. :0.000e+00	Min. :0.0000000	
## 1st Qu.:0.000000	1st Qu.:0.000e+00	1st Qu.:0.0000000	
## Median :0.000000	Median :0.000e+00	Median :0.0000000	
## Mean :0.001413	Mean :4.652e-05	Mean :0.0003517	
## 3rd Qu.:0.000000	3rd Qu.:0.000e+00	3rd Qu.:0.0000000	
## Max. :0.516477	Max. :1.682e-02	Max. :0.1274343	
## ris_facility_unknown	ris_mines_oilsands	ris_oilsands_rms	
## Min. :0.000000	Min. :0.0000000	Min. :0.000e+00	
## 1st Qu.:0.000000	1st Qu.:0.0000000	1st Qu.:0.000e+00	
## Median :0.000000	Median :0.0000000	Median :0.000e+00	
## Mean :0.000394	Mean :0.0000321	Mean :9.084e-05	
## 3rd Qu.:0.000000	3rd Qu.:0.0000000	3rd Qu.:0.000e+00	
## Max. :0.262453	Max. :0.0566685	Max. :3.350e-02	
## ris_overburden_dump	ris_reclaimed_permanent	ris_reclaimed_temp	
## Min. :0.0000000	Min. :0.0000000	Min. :0.000000	
## 1st Qu.:0.0000000	1st Qu.:0.0000000	1st Qu.:0.000000	
## Median :0.0000000	Median :0.0000000	Median :0.000000	
## Mean :0.0000473	Mean :0.0002917	Mean :0.000353	
## 3rd Qu.:0.0000000	3rd Qu.:0.0000000	3rd Qu.:0.000000	
## Max. :0.0211145	Max. :0.0534939	Max. :0.182379	
## ris_road	ris_soil_replaced	ris_soil_salvaged	
## Min. :0.0000000	Min. :0.000e+00	Min. :0.000e+00	
## 1st Qu.:0.0000000	1st Qu.:0.000e+00	1st Qu.:0.000e+00	
## Median :0.0000000	Median :0.000e+00	Median :0.000e+00	
## Mean :0.0001171	Mean :7.383e-05	Mean :7.075e-05	
## 3rd Qu.:0.0000000	3rd Qu.:0.000e+00	3rd Qu.:0.000e+00	
## Max. :0.0519727	Max. :2.448e-02	Max. :4.148e-02	
## ris_tailing_pond	ris_transmission_line	ris_utilities	
## Min. :0.0000000	Min. :0.000e+00	Min. :0.000e+00	
## 1st Qu.:0.0000000	1st Qu.:0.000e+00	1st Qu.:0.000e+00	
## Median :0.0000000	Median :0.000e+00	Median :0.000e+00	
## Mean :0.0003771	Mean :3.215e-06	Mean :2.503e-06	
## 3rd Qu.:0.0000000	3rd Qu.:0.000e+00	3rd Qu.:0.000e+00	
## Max. :0.1738171	Max. :2.667e-03	Max. :2.539e-03	
## ris_windrow	rlwy_sgl_track	road_gravel_1l	
## Min. :0.000e+00	Min. :0.000e+00	Min. :0.0000000	
## 1st Qu.:0.000e+00	1st Qu.:0.000e+00	1st Qu.:0.0000000	
## Median :0.000e+00	Median :0.000e+00	Median :0.0007062	

## Mean :1.099e-05	Mean :1.914e-05	Mean :0.0025547	
## 3rd Qu.:0.000e+00	3rd Qu.:0.000e+00	3rd Qu.:0.0036543	
## Max. :1.595e-02	Max. :4.500e-02	Max. :0.1021033	
## road_gravel_2l	road_paved_div	road_paved_undiv_1l	
## Min. :0.0000000	Min. :0.000e+00	Min. :0.000e+00	
## 1st Qu.:0.0000000	1st Qu.:0.000e+00	1st Qu.:0.000e+00	
## Median :0.0000000	Median :0.000e+00	Median :0.000e+00	
## Mean :0.0011781	Mean :2.219e-06	Mean :3.716e-05	
## 3rd Qu.:0.0004341	3rd Qu.:0.000e+00	3rd Qu.:0.000e+00	
## Max. :0.0438815	Max. :1.936e-03	Max. :2.147e-02	
## road_paved_undiv_2l	road_unclassified	road_unimproved	
## Min. :0.0000000	Min. :0.000e+00	Min. :0.0000000	
## 1st Qu.:0.0000000	1st Qu.:0.000e+00	1st Qu.:0.0000000	
## Median :0.0000000	Median :0.000e+00	Median :0.0001181	
## Mean :0.0003454	Mean :3.983e-06	Mean :0.0012657	
## 3rd Qu.:0.0000000	3rd Qu.:0.000e+00	3rd Qu.:0.0012586	
## Max. :0.0431663	Max. :1.027e-03	Max. :0.1216609	
## road_winter	rough_pasture	runway	
## Min. :0.000e+00	Min. :0.0000000	Min. :0.0000000	
## 1st Qu.:0.000e+00	1st Qu.:0.0000000	1st Qu.:0.0000000	
## Median :0.000e+00	Median :0.0000000	Median :0.0000000	
## Mean :7.283e-05	Mean :0.0001251	Mean :0.0001301	
## 3rd Qu.:0.000e+00	3rd Qu.:0.0000000	3rd Qu.:0.0000000	
## Max. :2.227e-02	Max. :0.0828324	Max. :0.0964794	
## rural_residence	sump	surrounding_veg	
## Min. :0.000e+00	Min. :0.0000000	Min. :0.0000000	
## 1st Qu.:0.000e+00	1st Qu.:0.0000000	1st Qu.:0.0000000	
## Median :0.000e+00	Median :0.0000000	Median :0.0000000	
## Mean :2.721e-05	Mean :0.0019291	Mean :0.0009727	
## 3rd Qu.:0.000e+00	3rd Qu.:0.0009333	3rd Qu.:0.0000000	
## Max. :2.805e-02	Max. :0.5090256	Max. :0.3339809	
## tailing_pond	tame_pasture	trail	transmission_line
## Min. :0.000e+00	Min. :0.0000000	Min. :0.000e+00	Min. :0.000000
## 1st Qu.:0.000e+00	1st Qu.:0.0000000	1st Qu.:5.325e-05	1st Qu.:0.000000
## Median :0.000e+00	Median :0.0000000	Median :6.083e-04	Median :0.000000
## Mean :9.661e-06	Mean :0.0004037	Mean :1.163e-03	Mean :0.002814
## 3rd Qu.:0.000e+00	3rd Qu.:0.0000000	3rd Qu.:1.491e-03	3rd Qu.:0.000000
## Max. :4.008e-03	Max. :0.1636895	Max. :8.091e-02	Max. :0.415142
## truck_trail	urban_industrial	urban_residence	
## Min. :0.0000000	Min. :0.0000000	Min. :0.000e+00	
## 1st Qu.:0.0000000	1st Qu.:0.0000000	1st Qu.:0.000e+00	
## Median :0.0000000	Median :0.0000000	Median :0.000e+00	
## Mean :0.0008846	Mean :0.0006373	Mean :2.019e-05	
## 3rd Qu.:0.0005300	3rd Qu.:0.0000000	3rd Qu.:0.000e+00	
## Max. :0.2417074	Max. :0.3357490	Max. :1.157e-02	
## vegetated_edge_railways	vegetated_edge_roads	well_aband	
## Min. :0.000e+00	Min. :0.000000	Min. :0.0000000	
## 1st Qu.:0.000e+00	1st Qu.:0.000000	1st Qu.:0.0001604	
## Median :0.000e+00	Median :0.006215	Median :0.0018274	
## Mean :8.062e-05	Mean :0.009250	Mean :0.0057321	
## 3rd Qu.:0.000e+00	3rd Qu.:0.012129	3rd Qu.:0.0064984	
## Max. :2.415e-01	Max. :0.168734	Max. :0.5861263	
## well_bitumen	well_cased	well_gas	
## Min. :0.0000000	Min. :0.0000000	Min. :0.0000000	


```

## 1st Qu.:0.0000000 1st Qu.:0.0000000 1st Qu.:0.0000000
## Median :0.0000000 Median :0.0000000 Median :0.0000000
## Mean :0.0037951 Mean :0.0003802 Mean :0.0003355
## 3rd Qu.:0.0002634 3rd Qu.:0.0000000 3rd Qu.:0.0000000
## Max. :0.2730052 Max. :0.1037580 Max. :0.0642192
## well_oil well_other well_unknown
## Min. :0.0000000 Min. :0.0000000 Min. :0.000e+00
## 1st Qu.:0.0000000 1st Qu.:0.0000000 1st Qu.:0.000e+00
## Median :0.0000000 Median :0.0000000 Median :0.000e+00
## Mean :0.0001851 Mean :0.0011533 Mean :1.613e-05
## 3rd Qu.:0.0000000 3rd Qu.:0.0005069 3rd Qu.:0.000e+00
## Max. :0.0672230 Max. :0.3153600 Max. :4.813e-03
## well_cleared_not_confirmed well_cleared_not_drilled array_year
## Min. :0.0000000 Min. :0.000e+00 Min. :2021
## 1st Qu.:0.0000000 1st Qu.:0.000e+00 1st Qu.:2022
## Median :0.0000000 Median :0.000e+00 Median :2022
## Mean :0.0001572 Mean :1.056e-05 Mean :2022
## 3rd Qu.:0.0000000 3rd Qu.:0.000e+00 3rd Qu.:2023
## Max. :0.1156142 Max. :1.469e-02 Max. :2023

```

```
names(hfi)
```

```

## [1] "array" "site"
## [3] "buffer_dist" "airp_runway"
## [5] "borrowpit_dry" "borrowpit_wet"
## [7] "borrowpits" "camp_industrial"
## [9] "campground" "canal"
## [11] "cfo" "clearing_unknown"
## [13] "clearing_wellpad_unconfirmed" "conventional_seismic"
## [15] "country_residence" "crop"
## [17] "cultivation_abandoned" "dugout"
## [19] "facility_other" "facility_unknown"
## [21] "greenspace" "grvl_sand_pit"
## [23] "harvest_area_1940" "harvest_area_1950"
## [25] "harvest_area_1960" "harvest_area_1963"
## [27] "harvest_area_1970" "harvest_area_1971"
## [29] "harvest_area_1972" "harvest_area_1973"
## [31] "harvest_area_1975" "harvest_area_1976"
## [33] "harvest_area_1977" "harvest_area_1978"
## [35] "harvest_area_1979" "harvest_area_1980"
## [37] "harvest_area_1984" "harvest_area_1985"
## [39] "harvest_area_1986" "harvest_area_1987"
## [41] "harvest_area_1988" "harvest_area_1989"
## [43] "harvest_area_1990" "harvest_area_1991"
## [45] "harvest_area_1992" "harvest_area_1993"
## [47] "harvest_area_1994" "harvest_area_1995"
## [49] "harvest_area_1996" "harvest_area_1997"
## [51] "harvest_area_1998" "harvest_area_1999"
## [53] "harvest_area_2000" "harvest_area_2001"
## [55] "harvest_area_2002" "harvest_area_2003"
## [57] "harvest_area_2004" "harvest_area_2005"
## [59] "harvest_area_2006" "harvest_area_2007"
## [61] "harvest_area_2008" "harvest_area_2009"

```

```
## [63] "harvest_area_2010"      "harvest_area_2011"
## [65] "harvest_area_2012"      "harvest_area_2013"
## [67] "harvest_area_2014"      "harvest_area_2015"
## [69] "harvest_area_2016"      "harvest_area_2017"
## [71] "harvest_area_2018"      "harvest_area_2019"
## [73] "harvest_area_2020"      "harvest_area_2021"
## [75] "harvest_area_white_zone" "lagoon"
## [77] "low_impact_seismic"     "mines_oilsands"
## [79] "misc_oil_gas_facility"  "oil_gas_plant"
## [81] "open_pit_mine"          "pipeline"
## [83] "reservoir"              "residence_clearing"
## [85] "ris_borrowpits"          "ris_camp_industrial"
## [87] "ris_clearing_unknown"   "ris_drainage"
## [89] "ris_facility_operations" "ris_facility_unknown"
## [91] "ris_mines_oilsands"     "ris_oilsands_rms"
## [93] "ris_overburden_dump"    "ris_reclaimed_permanent"
## [95] "ris_reclaimed_temp"     "ris_road"
## [97] "ris_soil_replaced"      "ris_soil_salvaged"
## [99] "ris_tailing_pond"       "ris_transmission_line"
## [101] "ris_utilities"          "ris_windrow"
## [103] "rlwy_sgl_track"         "road_gravel_1l"
## [105] "road_gravel_2l"         "road_paved_div"
## [107] "road_paved_undiv_1l"    "road_paved_undiv_2l"
## [109] "road_unclassified"      "road_unimproved"
## [111] "road_winter"            "rough_pasture"
## [113] "runway"                 "rural_residence"
## [115] "sump"                   "surrounding_veg"
## [117] "tailing_pond"           "tame_pasture"
## [119] "trail"                  "transmission_line"
## [121] "truck_trail"            "urban_industrial"
## [123] "urban_residence"        "vegetated_edge_railways"
## [125] "vegetated_edge_roads"   "well_aband"
## [127] "well_bitumen"           "well_cased"
## [129] "well_gas"               "well_oil"
## [131] "well_other"             "well_unknown"
## [133] "well_cleared_not_confirmed" "well_cleared_not_drilled"
## [135] "array_year"
```

1.2. Group the HFI data

We will follow the convention used by Marissa Dyck, ABMI, etc. to group variables. This code is nice and robust to whether a column is present in the data or not.

```
hfi_grouped <- hfi %>%

  # rename 'vegetated_edge_roads' so that we can use 'road' as keyword to group roads without
  # including this feature
  rename(vegetated_edge_rds = vegetated_edge_roads) %>%

  mutate(
    # borrowpits
    borrowpits = rowSums(across(any_of(c(
      grep("borrowpit", names(.), value = TRUE),
      'dugout',
```

```

    'lagoon',
    'sump'
  ))),

  # clearings
  clearings = rowSums(across(any_of(c(
    grep("clearing", names(.), value = TRUE),
    'runway'
  )))),

  # cultivations
  cultivation = rowSums(across(any_of(c(
    'crop',
    'cultivation_abandoned',
    'fruit_vegetables',
    'rough_pasture',
    'tame_pasture'
  )))),

  # industrial facilities
  facilities = rowSums(across(any_of(c(
    grep("facility", names(.), value = TRUE),
    grep("plant", names(.), value = TRUE),
    'camp_industrial',
    'urban_industrial',
    'mill',
    'ris_camp_industrial',
    'ris_tank_farm',
    'ris_utilities'
  )))),

  # harvest areas
  harvest_total = rowSums(across(contains('harvest'))),

  # mine areas
  mines = rowSums(across(any_of(c(
    grep("mine", names(.), value = TRUE),
    grep("tailing", names(.), value = TRUE),
    'grvl_sand_pit',
    'peat',
    'ris_drainage',
    'ris_oilsands_rms',
    'ris_overburden_dump',
    'ris_reclaim_ready',
    'ris_soil_salvaged',
    'ris_waste'
  )))),

  # railways
  railways = rowSums(across(any_of(grep("rlwy", names(.), value = TRUE)))),

  # reclaimed areas
  reclaimed = rowSums(across(any_of(c(
    grep("reclaimed", names(.), value = TRUE),
    'ris_soil_replaced',
    'ris_windrow'
  )))),

  # recreation areas

```

```

recreation = rowSums(across(any_of(c(
  'campground',
  'golfcourse',
  'greenspace',
  'recreation'
)))),

# residential areas
residential = rowSums(across(any_of(c(
  'country_residence',
  'rural_residence',
  'urban_residence'
)))),

# roads
roads = rowSums(across(any_of(c(
  grep("road", names(.), value = TRUE),
  'transfer_station',
  'interchange_ramp',
  'airp_runway',
  'ris_airp_runway'
)))),

# seismic lines
seismic_lines = conventional_seismic,

# 3D seismic lines
seismic_lines_3D = low_impact_seismic,

# transmission lines
transmission_lines = rowSums(across(any_of(c(
  grep("transmission", names(.), value = TRUE)
)))),

# trails
trails = rowSums(across(any_of(c(
  grep("trail", names(.), value = TRUE)
)))),

# vegetated edges
veg_edges = rowSums(across(any_of(c(
  grep("vegetated", names(.), value = TRUE),
  'surrounding_veg'
)))),

# man-made water features
water = rowSums(across(any_of(c(
  'canal',
  'reservoir'
)))),

# well sites
wells_total = rowSums(across(any_of(c(
  grep("well", names(.), value = TRUE)
)))),

# inactive well sites
wells_inactive = rowSums(across(any_of(c(
  "well_aband"

```

```

))))),

wells_active = rowSums(across(any_of(c(
  'well_bitumen',
  'well_gas',
  'well_oil',
  'well_cased',
  "well_cleared_not_confirmed",
  "well_cleared_not_drilled",
  "well_cleared_drilled",
  "well_drilled_other",
  "ris_well",
  "well_other",
  "well_unknown"
)))),

# remove columns that were used to create new columns to tidy the data frame
.keep = 'unused'
) %>%

# reorder alphabetically except array, site and buff_dist
select(order(colnames(.))) %>%

# move the columns that aren't HFI features or landcover to the front
relocate(array, site, buffer_dist, array_year) %>%

# reorder variables so the veg data is after all the HFI data
relocate(starts_with('lc_class'), .after = wells_active)

```

1.3. Group the harvest data by year

```

# Now compute row-wise harvest_0_15 and harvest_gt_15
# Loop through the sample years. This is the most efficient code I could find.
hfi_harvest <- purrr::map_dfr(unique(hfi$array_year), ~{

  # Columns to sum for each year
  cols_0_15 <- paste0("harvest_area_", seq(.x - 14, .x))
  cols_gt_15 <- paste0("harvest_area_", seq(1900, .x - 15))

  hfi %>%
    filter(array_year == .x) %>%
    mutate(

      # Recent harvest
      harvest_0_15 = rowSums(across(any_of(
        cols_0_15
      ))),

      # Old harvest
      harvest_gt_15 = rowSums(across(any_of(
        cols_gt_15
      )))
    ) %>%

  # move the columns that aren't HFI features or landcover to the front
  select(array, site, buffer_dist, array_year, harvest_0_15, harvest_gt_15)
}

```

```
}  
)
```

1.4. Add the harvest and do some additional grouping

```
hfi_final <- hfi_grouped %>%  
  
  left_join(hfi_harvest, by = c("array", "site", "buffer_dist", "array_year")) %>%  
  
  mutate(  
    # create column osm_industrial following Marissa's convention  
    osm_industrial =  
      borrowpits +  
      clearings +  
      facilities +  
      mines,  
  
    # Lump pipeline and transmission line  
    pipe_trans =  
      pipeline +  
      transmission_lines,  
  
    # Lump seismic  
    seismic =  
      seismic_lines +  
      seismic_lines_3D,  
  
    # Cumulative site disturbance without harvest  
    cfi_site =  
      osm_industrial +  
      wells_total +  
      seismic +  
      pipe_trans +  
      trails +  
      roads +  
      railways,  
  
    # Cumulative site disturbance with veg edges  
    cfi_site_with_vegedges =  
      cfi_site +  
      veg_edges,  
  
    # Cumulative site disturbance with harvest  
    cfi_site_with_harvest =  
      cfi_site +  
      harvest_0_15  
  ) %>%  
  
  # remove other features we don't need  
  select(!any_of(c(  
    'cfo',  
    'cultivation',  
    'reclaimed',  
    'recreation',  
    'residential',
```

```

    'water',
    'borrowpits',
    'clearings',
    'facilities',
    'mines',
    'pipeline',
    'transmission_lines'
  ))) %>%

  relocate(array, site, buffer_dist, array_year, sort(names(.)))

names(hfi_final)

```

```

## [1] "array"          "site"           "buffer_dist"
## [4] "array_year"     "cfi_site"       "cfi_site_with_harvest"
## [7] "cfi_site_with_vegedges" "harvest_0_15"   "harvest_gt_15"
## [10] "harvest_total"  "osm_industrial" "pipe_trans"
## [13] "railways"       "roads"          "seismic"
## [16] "seismic_lines"  "seismic_lines_3D" "trails"
## [19] "veg_edges"      "wells_active"   "wells_inactive"
## [22] "wells_total"

```

Looks good!!

3. Tidy up the forest inventory data

3.1. Clean up and add sample year column

```

sbfi <- sbfi %>%

  set_names(tolower(names(.)) %>%
    str_replace_all(., pattern = c('-'), # provide the character pattern to look for (if
    ↪ you don't keep the \ it won't work)
    replacement = '_') %>% # what you want the pattern to be replaced
    ↪ with
    str_replace_all(., pattern = c('\\.'), # provide the character pattern to look for
    ↪ (if you don't keep the \ it won't work)
    replacement = '_')
  ) %>%

  mutate(
    array_year = case_when(
      array %in% c("LU2", "LU3") ~ 2021,
      array %in% c("LU13", "LU15", "LU21", "LU1") ~ 2022,
      array %in% c("LU9", "LU16", "LU14", "LU22") ~ 2023
    )
  )

names(sbfi)

```

```

## [1] "array"          "site"           "buffer_dist"
## [4] "age_0_10"       "age_10_20"      "age_20_30"
## [7] "age_30_40"      "age_40_50"      "age_50_60"

```

```
## [10] "age_60_70"           "age_70_80"           "age_80_90"
## [13] "age_90_100"          "age_100_110"         "age_110_120"
## [16] "age_120_130"         "age_130_140"         "age_140_150"
## [19] "age_gt_150"          "lc_water"             "lc_snow_ice"
## [22] "lc_rock_rubble"      "lc_exposed_barren"    "lc_bryoids"
## [25] "lc_shrubs"           "lc_wetland"           "lc_wetland_treed"
## [28] "lc_herbs"            "lc_coniferous"        "lc_broadleaf"
## [31] "lc_mixedwood"        "pinu_ban_pct_of_treed" "popu_tre_pct_of_treed"
## [34] "pice_mar_pct_of_treed" "lari_lar_pct_of_treed" "pice_gla_pct_of_treed"
## [37] "betu_pap_pct_of_treed" "fire_pct_1985"        "fire_pct_1986"
## [40] "fire_pct_1987"        "fire_pct_1988"        "fire_pct_1989"
## [43] "fire_pct_1990"        "fire_pct_1991"        "fire_pct_1992"
## [46] "fire_pct_1993"        "fire_pct_1994"        "fire_pct_1995"
## [49] "fire_pct_1996"        "fire_pct_1997"        "fire_pct_1998"
## [52] "fire_pct_1999"        "fire_pct_2000"        "fire_pct_2001"
## [55] "fire_pct_2002"        "fire_pct_2003"        "fire_pct_2004"
## [58] "fire_pct_2005"        "fire_pct_2006"        "fire_pct_2007"
## [61] "fire_pct_2008"        "fire_pct_2009"        "fire_pct_2010"
## [64] "fire_pct_2011"        "fire_pct_2012"        "fire_pct_2013"
## [67] "fire_pct_2014"        "fire_pct_2015"        "fire_pct_2016"
## [70] "fire_pct_2017"        "fire_pct_2018"        "fire_pct_2019"
## [73] "fire_pct_2020"        "harvest_pct_1985"     "harvest_pct_1986"
## [76] "harvest_pct_1987"     "harvest_pct_1988"     "harvest_pct_1989"
## [79] "harvest_pct_1990"     "harvest_pct_1991"     "harvest_pct_1992"
## [82] "harvest_pct_1993"     "harvest_pct_1994"     "harvest_pct_1995"
## [85] "harvest_pct_1996"     "harvest_pct_1997"     "harvest_pct_1998"
## [88] "harvest_pct_1999"     "harvest_pct_2000"     "harvest_pct_2001"
## [91] "harvest_pct_2002"     "harvest_pct_2003"     "harvest_pct_2004"
## [94] "harvest_pct_2005"     "harvest_pct_2006"     "harvest_pct_2007"
## [97] "harvest_pct_2008"     "harvest_pct_2009"     "harvest_pct_2010"
## [100] "harvest_pct_2011"     "harvest_pct_2012"     "harvest_pct_2013"
## [103] "harvest_pct_2014"     "harvest_pct_2015"     "harvest_pct_2016"
## [106] "harvest_pct_2017"     "harvest_pct_2018"     "harvest_pct_2019"
## [109] "harvest_pct_2020"     "array_year"
```

3.2. Calculate fire by year columns

```
# Now compute row-wise fire_0_15 and fire_gt_15
# Loop through the sample years. This is the most efficient code I could find.
sbfi_fire <- purrr::map_dfr(unique(sbfi$array_year), ~{

  # Columns to sum (based on fire year and sample year)
  cols_0_15 <- paste0("fire_pct_", seq(.x - 14, .x))
  cols_gt_15 <- paste0("fire_pct_", seq(1900, .x - 15))

  sbfi %>%
    filter(array_year == .x) %>%
    mutate(

      # Sum them
      fire_0_15 = rowSums(across(any_of(
        cols_0_15
      ))),
```



```

    fire_gt_15 = rowSums(across(any_of(
      cols_gt_15
    )))
  ) %>%

# move the columns that aren't HFI features or landcover to the front
  select(array, site, buffer_dist, array_year, fire_0_15, fire_gt_15)
}
)

```

3.3. Finalize and give the species a more meaningful name

```

sbfi_final <- sbfi %>%

# Some columns we don't want
  select(-contains("fire"),
    -contains("age"),
    -lc_rock_rubble,
    -lc_snow_ice,
    -lc_bryoids,
    -lc_exposed_barren,
    -contains("harvest")) %>%

# Join the fire data
  left_join(sbfi_fire, by = c("array", "site", "buffer_dist", "array_year")) %>%

  relocate(array, site, buffer_dist, array_year, sort(names(.))) %>%

# Nicer species column names
  rename_with(
    ~ paste0("pct_", str_remove(., "_pct_of_treed")),
    .cols = contains("_pct_of_treed")
  )

names(sbfi_final)

```

```

## [1] "array"      "site"      "buffer_dist" "array_year"
## [5] "pct_betu_pap" "fire_0_15" "fire_gt_15" "pct_lari_lar"
## [9] "lc_broadleaf" "lc_coniferous" "lc_herbs" "lc_mixedwood"
## [13] "lc_shrubs" "lc_water" "lc_wetland" "lc_wetland_treed"
## [17] "pct_pice_gla" "pct_pice_mar" "pct_pinu_ban" "pct_popu_tre"

```

4. Format the configuration data

4.1. Merge data and select the metrics we want

```

config_final <- config1 %>%

  left_join(config2, by = c("site", "buffer")) %>%

```

```

# Classes for which we don't care about comp
select(-contains("water"),
       -contains("anthropogenic")
       ) %>%

# Tidy up the column names
rename_with(
  ~ paste0("nonanthro_", str_remove(., "natural_")),
  .cols = contains("natural_")
) %>%

rename(buffer_dist = buffer) %>%

# Fill in NA values with values that are "meaningful" for what we have.
# Check the landscapemetrics documentation to see what an NA should be
# An NA cannot meaningfully be replaced for all metrics (e.g., 0/0)
mutate(
  across(contains("_tca"), ~ ifelse(is.na(.), 0, .)),
  across(contains("_ed"), ~ ifelse(is.na(.), 0, .)),
  across(contains("_cohesion"), ~ ifelse(is.na(.), 100, .)),
  across(contains("_np"), ~ ifelse(is.na(.), 1, .)),
  across(contains("_cai_mn"), ~ ifelse(is.na(.), 0, .)),
  across(contains("_siei"), ~ ifelse(is.na(.), 0.001, .)),
  across(contains("_contag"), ~ ifelse(is.na(.), max(., na.rm=T), .))
)

summary(config_final)

```

```

##   buffer_dist      site      nonanthro_cai_mn nonanthro_ed
##   Min.   : 50      Length:9460      Min.   : 0.00      Min.   : 0.00
##   1st Qu.:1000     Class :character      1st Qu.:20.67     1st Qu.: 48.55
##   Median :2375     Mode  :character      Median :32.73     Median : 74.60
##   Mean   :2393                                     Mean   :36.34     Mean   :101.65
##   3rd Qu.:3750                                     3rd Qu.:49.51     3rd Qu.:123.53
##   Max.   :5000                                     Max.   :98.21     Max.   :742.03
##   nonanthro_tca    landscape_cai_mn landscape_ed landscape_tca
##   Min.   : 0.0      Min.   :0      Min.   :0      Min.   :0
##   1st Qu.: 157.3    1st Qu.:0      1st Qu.:0      1st Qu.:0
##   Median :1024.5    Median :0      Median :0      Median :0
##   Mean   :1664.3    Mean   :0      Mean   :0      Mean   :0
##   3rd Qu.:2766.3    3rd Qu.:0      3rd Qu.:0      3rd Qu.:0
##   Max.   :7585.5    Max.   :0      Max.   :0      Max.   :0
##   nonanthro_cohesion landscape_cohesion landscape_contag landscape_mesh
##   Min.   : 36.33    Min.   : 90.32    Min.   :18.38    Min.   : 0.1503
##   1st Qu.: 99.17    1st Qu.: 99.24    1st Qu.:57.72    1st Qu.: 36.4937
##   Median : 99.57    Median : 99.51    Median :62.84    Median : 113.0812
##   Mean   : 99.22    Mean   : 99.26    Mean   :63.54    Mean   : 192.7786
##   3rd Qu.: 99.79    3rd Qu.: 99.65    3rd Qu.:68.62    3rd Qu.: 242.2693
##   Max.   :100.00    Max.   :100.00    Max.   :99.53    Max.   :2978.4634
##   landscape_np      landscape_shei landscape_siei
##   Min.   : 1.0      Min.   :0.0000    Min.   :0.0010
##   1st Qu.: 72.0     1st Qu.:0.5421    1st Qu.:0.6306
##   Median : 340.0     Median :0.6390    Median :0.7549
##   Mean   : 809.8     Mean   :0.6216    Mean   :0.7005
##   3rd Qu.: 939.2     3rd Qu.:0.7252    3rd Qu.:0.8268

```

```
## Max.      :16279.0    Max.      :1.0000    Max.      :1.0000
```

5. Put it all together:

5.1. Join the data

```
covs <- hfi_final %>%  
  left_join(sbf_i_final, by = c("array", "site", "buffer_dist", "array_year")) %>%  
  left_join(config_final, by = c("site", "buffer_dist"))
```

5.2. Append lat/long and UTM 12N coordinates

```
coords <- read_csv("../data/raw/OSM_coordinates_2021_2022_2023.csv")  
  
## Rows: 433 Columns: 4  
## -- Column specification -----  
## Delimiter: ","  
## chr (2): array, site  
## dbl (2): lat, long  
##  
## 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.  
  
# assuming lat/long was extracted in NAD83  
coords <- st_as_sf(coords, coords = c("long", "lat"), crs = 4269) %>%  
  
  select(array, site) %>%  
  
  # Get lat and long in epsg:4269 in their own columns  
  mutate(long = st_coordinates(.)[,1],  
         lat = st_coordinates(.)[,2]) %>%  
  
  # NAD1983 UTM 12N  
  st_transform(osm_coords, crs = 26912) %>%  
  
  # Get UTM in columns  
  mutate(easting_12n = st_coordinates(.)[,1],  
         northing_12n = st_coordinates(.)[,2]) %>%  
  
  st_drop_geometry(.)
```

5.3. Final inspection

```
# Tidy up and final check
```

```
covs_final <- covs %>%
```

```
left_join(coords, by = c('array', 'site')) %>%
```

```
relocate(array, site, array_year, lat, long, easting_12n, northing_12n, buffer_dist)
```

```
summary(covs_final)
```

```
##      array      site      array_year      lat
## Length:9460    Length:9460    Min.   :2021    Min.   :54.56
## Class :character Class :character 1st Qu.:2022    1st Qu.:55.54
## Mode  :character Mode  :character Median :2022    Median :56.55
##                                     Mean  :2022    Mean   :56.30
##                                     3rd Qu.:2023    3rd Qu.:57.09
##                                     Max.   :2023    Max.   :57.56
##      long      easting_12n    northing_12n    buffer_dist
## Min.   :-115.0    Min.   :254793    Min.   :6046183    Min.   : 50
## 1st Qu.: -113.0    1st Qu.:374389    1st Qu.:6155043    1st Qu.:1000
## Median : -111.8    Median :448550    Median :6267631    Median :2375
## Mean   : -112.2    Mean   :426568    Mean   :6241165    Mean   :2393
## 3rd Qu.: -111.2    3rd Qu.:487062    3rd Qu.:6328625    3rd Qu.:3750
## Max.   : -110.1    Max.   :559033    Max.   :6379755    Max.   :5000
##      cfi_site      cfi_site_with_harvest cfi_site_with_vegedges
## Min.   :0.00000    Min.   :0.00000    Min.   :0.00000
## 1st Qu.:0.01940    1st Qu.:0.02623    1st Qu.:0.02479
## Median :0.04649    Median :0.06117    Median :0.05614
## Mean   :0.07117    Mean   :0.08914    Mean   :0.08147
## 3rd Qu.:0.09009    3rd Qu.:0.11801    3rd Qu.:0.10414
## Max.   :1.00000    Max.   :1.00000    Max.   :1.00000
##      harvest_0_15    harvest_gt_15    harvest_total    osm_industrial
## Min.   :0.00000    Min.   :0.00000    Min.   :0.00000    Min.   :0.0000000
## 1st Qu.:0.00000    1st Qu.:0.00000    1st Qu.:0.00000    1st Qu.:0.0000695
## Median :0.00000    Median :0.00000    Median :0.00000    Median :0.0036825
## Mean   :0.01797    Mean   :0.03621    Mean   :0.05430    Mean   :0.0205346
## 3rd Qu.:0.00000    3rd Qu.:0.02661    3rd Qu.:0.04944    3rd Qu.:0.0176253
## Max.   :1.00000    Max.   :1.00000    Max.   :1.00000    Max.   :1.0000000
##      pipe_trans      railways      roads      seismic
## Min.   :0.000000    Min.   :0.000e+00    Min.   :0.0000000    Min.   :0.000000
## 1st Qu.:0.000000    1st Qu.:0.000e+00    1st Qu.:0.0004244    1st Qu.:0.003467
## Median :0.007815    Median :0.000e+00    Median :0.0040688    Median :0.007971
## Mean   :0.017861    Mean   :1.914e-05    Mean   :0.0056152    Mean   :0.012823
## 3rd Qu.:0.022839    3rd Qu.:0.000e+00    3rd Qu.:0.0076654    3rd Qu.:0.015868
## Max.   :0.462911    Max.   :4.500e-02    Max.   :0.1216609    Max.   :0.199495
##      seismic_lines    seismic_lines_3D      trails      veg_edges
## Min.   :0.000000    Min.   :0.0000000    Min.   :0.0000000    Min.   :0.000000
## 1st Qu.:0.002977    1st Qu.:0.0000000    1st Qu.:0.0001309    1st Qu.:0.000000
## Median :0.006318    Median :0.0000000    Median :0.0010088    Median :0.006269
## Mean   :0.008086    Mean   :0.0047371    Mean   :0.0020477    Mean   :0.010303
## 3rd Qu.:0.009888    3rd Qu.:0.0003958    3rd Qu.:0.0024323    3rd Qu.:0.012470
## Max.   :0.151463    Max.   :0.1144061    Max.   :0.2417074    Max.   :0.359208
##      wells_active    wells_inactive    wells_total
## Min.   :0.0000000    Min.   :0.0000000    Min.   :0.0000000
```

```

## 1st Qu.:0.0000000 1st Qu.:0.0001604 1st Qu.:0.0006221
## Median :0.0004342 Median :0.0018274 Median :0.0045921
## Mean :0.0060331 Mean :0.0057321 Mean :0.0122699
## 3rd Qu.:0.0053666 3rd Qu.:0.0064984 3rd Qu.:0.0138036
## Max. :0.3153600 Max. :0.5861263 Max. :0.5861263
## pct_betu_pap fire_0_15 fire_gt_15 pct_lari_lar
## Min. :0.000e+00 Min. :0.0000000 Min. :0.0000000 Min. :0.000000
## 1st Qu.:0.000e+00 1st Qu.:0.0000000 1st Qu.:0.0000000 1st Qu.:0.000000
## Median :0.000e+00 Median :0.0000000 Median :0.0000000 Median :0.00421
## Mean :1.474e-07 Mean :0.0693643 Mean :0.036241 Mean :0.02216
## 3rd Qu.:0.000e+00 3rd Qu.:0.0009605 3rd Qu.:0.002022 3rd Qu.:0.02925
## Max. :1.640e-04 Max. :1.0000000 Max. :1.0000000 Max. :0.46693
## lc_broadleaf lc_coniferous lc_herbs lc_mixedwood
## Min. :0.00000 Min. :0.0000 Min. :0.0000000 Min. :0.00000
## 1st Qu.:0.02089 1st Qu.:0.2376 1st Qu.:0.002124 1st Qu.:0.01491
## Median :0.07534 Median :0.3410 Median :0.013893 Median :0.04786
## Mean :0.14876 Mean :0.3514 Mean :0.032364 Mean :0.06846
## 3rd Qu.:0.22015 3rd Qu.:0.4483 3rd Qu.:0.042228 3rd Qu.:0.09372
## Max. :0.99740 Max. :0.9997 Max. :0.431069 Max. :0.71213
## lc_shrubs lc_water lc_wetland lc_wetland_treed
## Min. :0.0000000 Min. :0.000000 Min. :0.00000 Min. :0.0000
## 1st Qu.:0.0000334 1st Qu.:0.000024 1st Qu.:0.01476 1st Qu.:0.1143
## Median :0.0005785 Median :0.001128 Median :0.03612 Median :0.2282
## Mean :0.0263576 Mean :0.016630 Mean :0.06793 Mean :0.2529
## 3rd Qu.:0.0036030 3rd Qu.:0.012358 3rd Qu.:0.09015 3rd Qu.:0.3595
## Max. :0.9556450 Max. :0.591750 Max. :0.76425 Max. :0.9508
## pct_pice_gla pct_pice_mar pct_pinu_ban pct_popu_tre
## Min. :0.0000000 Min. :0.0000 Min. :0.0000000 Min. :0.0000
## 1st Qu.:0.0000000 1st Qu.:0.2844 1st Qu.:0.001368 1st Qu.:0.1594
## Median :0.0000479 Median :0.5628 Median :0.006014 Median :0.3225
## Mean :0.0118011 Mean :0.5166 Mean :0.065547 Mean :0.3646
## 3rd Qu.:0.0085401 3rd Qu.:0.7394 3rd Qu.:0.031452 3rd Qu.:0.5568
## Max. :0.3872200 Max. :1.0000 Max. :1.0000000 Max. :1.0000
## nonanthro_cai_mn nonanthro_ed nonanthro_tca landscape_cai_mn
## Min. : 0.00 Min. : 0.00 Min. : 0.0 Min. :0
## 1st Qu.:20.67 1st Qu.: 48.55 1st Qu.: 157.3 1st Qu.:0
## Median :32.73 Median : 74.60 Median :1024.5 Median :0
## Mean :36.34 Mean :101.65 Mean :1664.3 Mean :0
## 3rd Qu.:49.51 3rd Qu.:123.53 3rd Qu.:2766.3 3rd Qu.:0
## Max. :98.21 Max. :742.03 Max. :7585.5 Max. :0
## landscape_ed landscape_tca nonanthro_cohesion landscape_cohesion
## Min. :0 Min. :0 Min. : 36.33 Min. : 90.32
## 1st Qu.:0 1st Qu.:0 1st Qu.: 99.17 1st Qu.: 99.24
## Median :0 Median :0 Median : 99.57 Median : 99.51
## Mean :0 Mean :0 Mean : 99.22 Mean : 99.26
## 3rd Qu.:0 3rd Qu.:0 3rd Qu.: 99.79 3rd Qu.: 99.65
## Max. :0 Max. :0 Max. :100.00 Max. :100.00
## landscape_contag landscape_mesh landscape_np landscape_shei
## Min. :18.38 Min. : 0.1503 Min. : 1.0 Min. :0.0000
## 1st Qu.:57.72 1st Qu.: 36.4937 1st Qu.: 72.0 1st Qu.:0.5421
## Median :62.84 Median : 113.0812 Median : 340.0 Median :0.6390
## Mean :63.54 Mean : 192.7786 Mean : 809.8 Mean :0.6216
## 3rd Qu.:68.62 3rd Qu.: 242.2693 3rd Qu.: 939.2 3rd Qu.:0.7252
## Max. :99.53 Max. :2978.4634 Max. :16279.0 Max. :1.0000

```

```
## landscape_siei
## Min.      :0.0010
## 1st Qu.:0.6306
## Median :0.7549
## Mean      :0.7005
## 3rd Qu.:0.8268
## Max.      :1.0000
```

5.4. Save the data:

```
write_csv(covs_final, "./data/processed/OSM_all_covariates_HFI_SBFI_final.csv")
```

End script