

# Living Data Project PREE Manuscript

## Spatial distribution of saskatoon (*Amelanchier alnifolia*) in central Saskatchewan

Hannah Pilat

2022-09-28

```
library(remotes)
remotes::install_github('Pakillo/grateful')
```

```
## Skipping install of 'grateful' from a github remote, the SHA1 (ba9b003e) has not changed since last
## Use `force = TRUE` to force installation
```

```
library(grateful)
get_pkgs_info()
```

```
##          pkg version                                citekeys
## 1      base    4.2.1                                base
## 2 grateful  0.1.11                                grateful
## 3 groundhog  2.0.1                                groundhog
## 4      knitr    1.40      knitr2022, knitr2015, knitr2014
## 5    remotes   2.4.2                                remotes
## 6 rmarkdown  2.16 rmarkdown2022, rmarkdown2018, rmarkdown2020
```

```
#code to prevent new directories being created if those directories already exist
if(!dir.exists("data")){dir.create("data")}

#create a folder for data
dir.create("data/")
```

```
14 ## Warning in dir.create("data/"): 'data' already exists
```

```
#create subfolders for raw and clean data within the data folder  
dir.create("data/raw")
```

```
15 ## Warning in dir.create("data/raw"): 'data/raw' already exists
```

```
dir.create("data/clean")
```

```
16 ## Warning in dir.create("data/clean"): 'data/clean' already exists
```

```
17 {knitr::opts_chunk$set(echo = TRUE)}
```

```
#create an object for the GBIF data for the species of interest  
saskatoon_occ <- ("Amelanchier alnifolia")
```

```
#download GBIF occurrence records within a specified window of coordinates:  
#list coordinates small first, large second  
gbif_data <- occ_data(scientificName = saskatoon_occ, hasCoordinate = TRUE,  
                      limit = 20000, decimalLongitude = "-110,-103",  
                      decimalLatitude = "52, 56")
```

```
#get a map of Saskatchewan from mapcan data by specifying the boundaries of the province  
sask_map <- mapcan(boundaries = province,  
                  province = SK,  
                  type = standard)
```

```
#get the columns that are relevant for mapping saskatoon's distribution  
#since these are presence data, we can get rid of the occurrence status column  
#select the longitude and latitude columns  
gbif_data_relevant <- gbif_data$data[ , c("decimalLongitude", "decimalLatitude")]  
  
#check the data to confirm only the relevant columns are included  
view(gbif_data_relevant) #or  
head(gbif_data_relevant)
```

```

18 ## # A tibble: 6 x 2
19 ##   decimalLongitude decimalLatitude
20 ##   <dbl>          <dbl>
21 ## 1    -107.         52.2
22 ## 2    -107.         52.1
23 ## 3    -107.         52.0
24 ## 4    -107.         52.0
25 ## 5    -109.         53.4
26 ## 6    -107.         52.2

```

```

#the relevant data are now in a dataframe, save it as a .csv file in the clean data folder
write.csv(gbif_data_relevant, file = "data/clean/gbif-saskatoon-occurrences.csv")

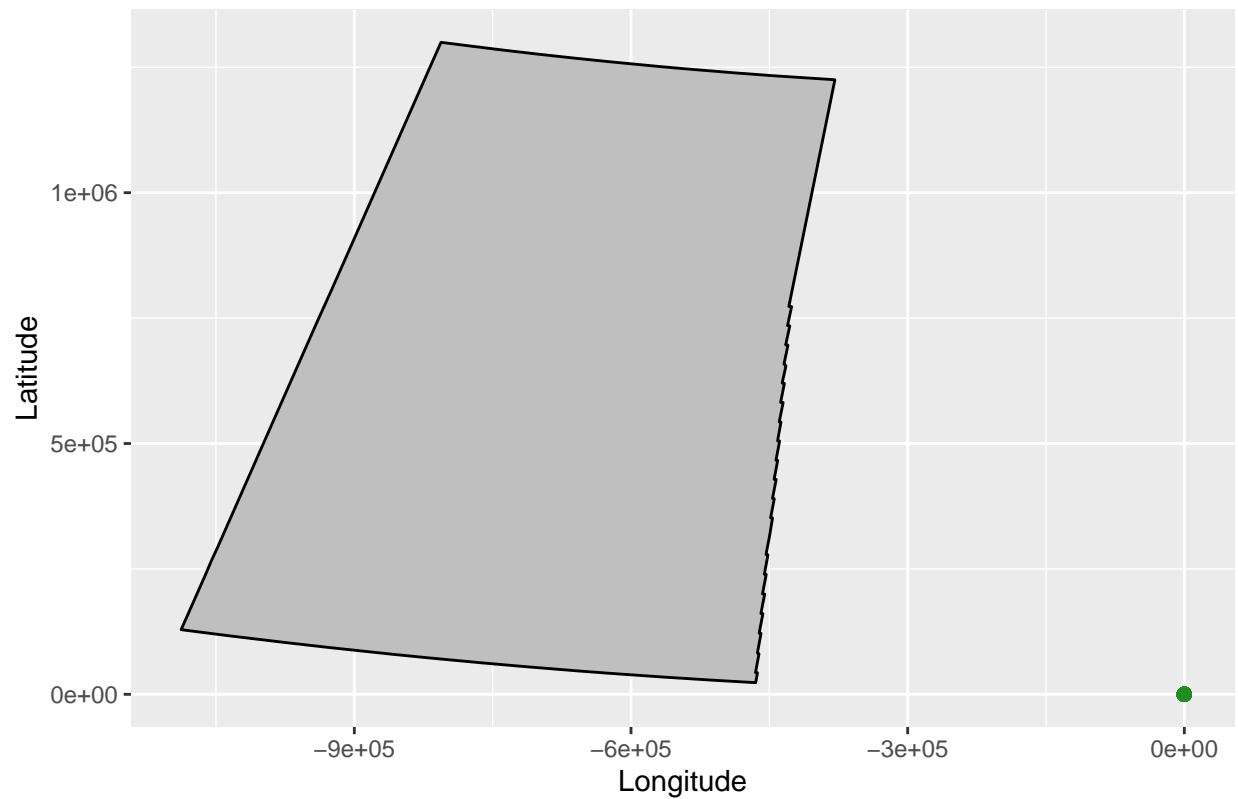
```

```

#plot the Saskatchewan data and create a polygon to outline and fill the province
ggplot(data = sask_map, mapping = aes(x = long, y = lat)) +
  geom_polygon(data = sask_map, mapping = aes(group = group), fill = "grey75",
    colour = "black", show.legend = FALSE) +
#next step is to plot the occurrences of saskatoon plants, but need to adjust the...
#coordinate reference system so it's consistent between sask_map and gbif_data_relevant data
  geom_jitter(data = gbif_data_relevant, aes(x = decimalLongitude, y = decimalLatitude),
    alpha = 0.5, size = 2, colour = "forestgreen", show.legend = FALSE) +
  labs(x = 'Longitude', y = 'Latitude') + #create more informative axes labels
  ggtitle("Amelanchier alnifolia occurrences in central Saskatchewan") #add a title to the figure

```

## Amelanchier alnifolia occurrences in central Saskatchewan



27

```
cite_packages(output = 'paragraph')
```

28

```
## We used R version 4.2.1 [@base] and the following R packages: grateful v. 0.1.11 [@grateful], ground
```

29

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated

30

the plot.