1 Data sets

1.1 Forest defoliation

The data collection is coordinated by the Thünen Institute (a research institute for forestry and agriculture). Every year approx 10,000 trees are surveyed across Germany at a regular 16 by 16 km grid. Several covariates are collected, but we will focus here on the mean defoliation per sampling point. Several trees are measured at each stand and averaged. For each sampling point we obtain fore each year and species one value.

A subset of the data are made available here(https://blumwald.thuenen.de/wze/aktuelle-ergebnisse-der-wze). For our purposes this subset is sufficient.

The data set was used in several other publication, that might be of interst:

- Eickenscheidt, Augustin, and Wellbrock (2018) used the data set to predict spatio-temporal patterns of tree defoliation.

1.1.1 How to read the data

```
dat <- read_rds("data/trees.rds")
dat
```

```
# A tibble: 66,409 x 29
    year point_id sp         mean_loss max_loss n_trees   ele slope aspect   TRI
   <dbl>    <dbl> <chr>          <dbl>    <dbl>   <dbl> <dbl> <dbl>  <dbl> <dbl>
 1  1990    10003 spruce          25         35      12    69  1.88   194.  22.6
 2  1990    10003 pine            10.6       30       8    69  1.88   194.  22.6
 3  1990    10003 deciduous~      21.2       25       4    69  1.88   194.  22.6
 4  1990    10003 coniferou~      19.2       35      20    69  1.88   194.  22.6
 5  1990    10003 beech           21.2       25       4    69  1.88   194.  22.6
 6  1990    10003 all speci~      19.6       35      24    69  1.88   194.  22.6
 7  1991    10003 spruce          25         35      12    69  1.88   194.  22.6
 8  1991    10003 pine            10.6       30       8    69  1.88   194.  22.6
 9  1991    10003 deciduous~      21.2       25       4    69  1.88   194.  22.6
```

```
10  1991    10003 coniferou~      19.2       35      20    69  1.88   194.  22.6
# i 66,399 more rows
# i 19 more variables: bio1 <dbl>, bio2 <dbl>, bio3 <dbl>, bio4 <dbl>,
#   bio5 <dbl>, bio6 <dbl>, bio7 <dbl>, bio8 <dbl>, bio9 <dbl>, bio10 <dbl>,
#   bio11 <dbl>, bio12 <dbl>, bio13 <dbl>, bio14 <dbl>, bio15 <dbl>,
#   bio16 <dbl>, bio17 <dbl>, bio18 <dbl>, bio19 <dbl>
```

There are more columns:

```
names(dat)
```

```
 [1] "year"      "point_id"  "sp"        "mean_loss" "max_loss"  "n_trees"
 [7] "ele"       "slope"     "aspect"    "TRI"       "bio1"      "bio2"
[13] "bio3"      "bio4"      "bio5"      "bio6"      "bio7"      "bio8"
[19] "bio9"      "bio10"     "bio11"     "bio12"     "bio13"     "bio14"
[25] "bio15"     "bio16"     "bio17"     "bio18"     "bio19"
```

TRI stands for terrain ruggedness index. The meaning of the bio covariates is:

- BIO1 = Annual Mean Temperature
- BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
- BIO3 = Isothermality (BIO2/BIO7) ($\times$100)
- BIO4 = Temperature Seasonality (standard deviation $\times$100)
- BIO5 = Max Temperature of Warmest Month
- BIO6 = Min Temperature of Coldest Month
- BIO7 = Temperature Annual Range (BIO5-BIO6)
- BIO8 = Mean Temperature of Wettest Quarter
- BIO9 = Mean Temperature of Driest Quarter
- BIO10 = Mean Temperature of Warmest Quarter
- BIO11 = Mean Temperature of Coldest Quarter
- BIO12 = Annual Precipitation
- BIO13 = Precipitation of Wettest Month
- BIO14 = Precipitation of Driest Month
- BIO15 = Precipitation Seasonality (Coefficient of Variation)
- BIO16 = Precipitation of Wettest Quarter
- BIO17 = Precipitation of Driest Quarter
- BIO18 = Precipitation of Warmest Quarter
- BIO19 = Precipitation of Coldest Quarter

There are observation for the following trees

```r
unique(dat$sp)
```

```
[1] "spruce"                "pine"                  "deciduous trees"
[4] "coniferous trees"      "beech"                 "all species"
[7] "other coniferous trees" "oak"                  "other deciduous trees"
```

1.2 Germany and German states

An outline for Germany and its states is saved here:

1.3 Climate data

Some climate data (not yet completed) are available from here: