data_monthly_v1_0_EDA

0. Preparation

Load Libraries

```
# Load libraries
library(tidyverse)
library(Hmisc)
library(gridExtra)
library(finalfit)
library(stargazer)
# To create and work with tidy temporal data
library(tsibble)
# To work with date-times and time-spans
library(lubridate)
# Provides a collection of commonly used univariate/multivariate TS models
library(fable)
## To interact directly with the Quandl API and download data
library(Quandl)
# For analyzing tidy time series data.
library(feasts)
# Provides methods and tools for displaying and analyzing univariate time series forecasts library(fore
# For estimation, lag selection, diagnostic testing, forecasting, and impulse response functions of VAR
#provides tools for statistical calculations
library(stats)
# To assist the quantitative trader in the development,
#testing and deployment of statistically based trading models.
library(quantmod)
# For statistical analysis
library(car)
## To retrieve and display the information returned online by Google Trends
library(gtrendsR)
# To do time series analysis and computational finance.
library(tseries)
```

Import data from csv file

Rows: 19015 Columns: 62

Link to the handmade codebook https://docs.google.com/spreadsheets/d/1-FQsF_sxnA6iBMNpHmGovkj9Xev6BCYg2j3Poj
edit?usp=sharing

Overview of monthly dataset

```
raw_df <- read_csv("../../../data/datasets/data_monthly_v1_0.csv")
glimpse(raw_df)
```

```
## Rows: 19,015
## Columns: 62
## $ SITE_ID           <chr> "AR-SLu", "AR-SLu", "AR-SLu", "AR-SLu", "AR-SLu", "A~
## $ year              <dbl> 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010~
## $ month             <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 1, 2, 3, 3, 4~
## $ TIMESTAMP         <dbl> 201001, 201002, 201003, 201004, 201005, 201006, 2010~
## $ dataset           <chr> "FLUXNET", "FLUXNET", "FLUXNET", "FLUXNET", "FLUXNET~
## $ SITE_IGBP         <chr> "MF", "MF", "MF", "MF", "MF", "MF", "MF", "MF", "MF"~
## $ LOCATION_LAT      <dbl> -33.4648, -33.4648, -33.4648, -33.4648, -33.4648, -3~
## $ LOCATION_LONG     <dbl> -66.4598, -66.4598, -66.4598, -66.4598, -66.4598, -6~
## $ TA_F              <dbl> 28.493, 26.673, 25.744, 18.450, 13.493, 11.273, 8.13~
## $ VPD_F             <dbl> 23.378, 14.369, 15.167, 9.185, 5.823, 5.219, 4.949, ~
## $ P_F               <dbl> 0.903, 1.986, 0.371, 0.100, 1.852, 0.030, 0.003, 0.1~
## $ NETRAD            <dbl> 188.59881, 144.21620, 125.64314, 71.50069, 41.24915,~
## $ NEE_VUT_REF       <dbl> -5.6327800, -4.4743300, -3.8928800, -3.1115900, -1.7~
## $ NEE_VUT_REF_QC    <dbl> 0.944892, 0.969494, 0.938844, 0.962500, 0.895833, 0.~
## $ NEE_CUT_REF       <dbl> -5.627700, -4.453580, -3.884050, -3.107050, -1.55985~
## $ NEE_CUT_REF_QC    <dbl> 0.948253, 0.970982, 0.938844, 0.962500, 0.913978, 0.~
## $ GPP_NT_VUT_REF    <dbl> 10.20950, 8.16307, 7.06222, 5.72781, 3.47763, 2.7738~
## $ GPP_DT_VUT_REF    <dbl> 11.91330, 9.97563, 9.00824, 6.54333, 4.15484, 3.7793~
## $ GPP_NT_CUT_REF    <dbl> 10.08900, 8.09051, 7.07681, 5.65260, 3.56473, 2.9719~
## $ GPP_DT_CUT_REF    <dbl> 11.92320, 10.16630, 9.00492, 6.60730, 4.16886, 3.691~
## $ RECO_NT_VUT_REF   <dbl> 4.46072, 3.62522, 3.18909, 2.55268, 1.78421, 2.86141~
## $ RECO_DT_VUT_REF   <dbl> 7.03163, 5.68557, 6.51721, 4.14082, 3.35165, 3.07628~
## $ RECO_NT_CUT_REF   <dbl> 4.45634, 3.61530, 3.18613, 2.54582, 1.84822, 2.83905~
## $ RECO_DT_CUT_REF   <dbl> 7.06081, 6.02964, 6.61985, 4.04346, 3.54627, 3.19315~
## $ time              <chr> "1/31/10", "2/28/10", "3/31/10", "4/30/10", "5/31/10~
## $ ET                <dbl> 9.014540, 7.677973, 5.890317, 2.345664, 2.208000, 1.~
## $ `BESS-PAR`        <dbl> 154, 120, 107, 81, 56, 48, 58, 72, 92, 124, 147, 157~
## $ `BESS-PARdiff`    <dbl> 40, 46, 31, 27, 19, 15, 15, 23, 30, 35, 39, 38, 42, ~
## $ `BESS-RSDN`       <dbl> 336, 258, 231, 175, 122, 105, 129, 158, 202, 271, 32~
## $ `CSIF-SIFdaily`   <dbl> 0.20432499, 0.14553030, 0.10980482, 0.07672890, 0.06~
## $ `CSIF-SIFinst`    <dbl> 0.5166268, 0.3872625, 0.3072417, 0.2238720, 0.200635~
## $ PET               <dbl> -0.013386652, -0.008937791, -0.008132122, -0.0067583~
## $ Ts                <dbl> 302.4697, 298.7886, 297.5482, 291.6960, 287.0565, 28~
## $ Tmean             <dbl> 300.1098, 297.2751, 296.4367, 290.6138, 286.8832, 28~
## $ prcp              <dbl> 0.002115019, 0.003131761, 0.002206154, 0.000209161, ~
## $ vpd               <dbl> 2.0661800, 1.0901145, 1.1686398, 0.9461956, 0.716290~
## $ `prcp-lag3`       <dbl> 0.008738702, 0.009724296, 0.007452934, 0.005547076, ~
## $ `ESACCI-sm`       <dbl> 0.1515208, 0.1665578, 0.1640767, 0.1240165, 0.142726~
## $ MODIS_LC          <dbl> 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 9, 9, 9~
## $ b1                <dbl> 0.08442580, 0.09180000, 0.08556129, 0.08740333, 0.07~
## $ b2                <dbl> 0.2687742, 0.2524464, 0.2304000, 0.2158133, 0.198193~
## $ b3                <dbl> 0.04531612, 0.04803214, 0.04453871, 0.04530000, 0.03~
## $ b4                <dbl> 0.08052903, 0.08092143, 0.07463870, 0.07454334, 0.06~
## $ b5                <dbl> 0.3005774, 0.2924500, 0.2694322, 0.2600900, 0.231483~
## $ b6                <dbl> 0.2505258, 0.2522143, 0.2411645, 0.2348300, 0.202816~
## $ b7                <dbl> 0.15535806, 0.15945713, 0.15363870, 0.14688666, 0.12~
```

```
## $ EVI               <dbl> 0.3212592, 0.2783001, 0.2568952, 0.2292145, 0.223529~
## $ GCI               <dbl> 2.349203, 2.121655, 2.087431, 1.895962, 2.061106, 1.~
## $ NDVI              <dbl> 0.5227052, 0.4668434, 0.4583453, 0.4235885, 0.445367~
## $ NDWI              <dbl> 0.035415366, 0.000400779, -0.022856813, -0.042024087~
## $ NIRv              <dbl> 0.14050612, 0.11781442, 0.10565167, 0.09139932, 0.08~
## $ kNDVI             <dbl> 0.26745087, 0.21459042, 0.20721813, 0.17754844, 0.19~
## $ Percent_Snow      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ Fpar              <dbl> 0.49, 0.43, 0.41, 0.36, 0.37, 0.33, 0.32, 0.26, 0.25~
## $ Lai               <dbl> 1.2, 0.9, 0.8, 0.5, 0.5, 0.4, 0.4, 0.3, 0.4, 0.6, 0.~
## $ LST_Day           <dbl> 313.84, 309.86, 309.18, 303.24, 296.20, 293.18, 292.~
## $ LST_Night         <dbl> 293.58, 292.96, 290.52, 286.34, 277.82, 276.80, 271.~
## $ MODIS_IGBP        <chr> "OSH", "OSH", "OSH", "OSH", "OSH", "OSH", "OSH", "OS~
## $ MODIS_PFT         <chr> "SH", "SH", "SH", "SH", "SH", "SH", "SH", "SH", "SH"~
## $ koppen_sub        <chr> "BSk", "BSk", "BSk", "BSk", "BSk", "BSk", "BSk", "BS~
## $ koppen            <chr> "Arid", "Arid", "Arid", "Arid", "Arid", "Arid", "Ari~
## $ CO2_concentration <dbl> 387.110, 387.675, 388.195, 388.905, 389.320, 389.160~
```

1. Findings

Source of each observation

All the observation seems to have origin in four datasets

```
raw_df %>%
  group_by(dataset) %>%
  summarise(count=n())
```

```
## # A tibble: 4 x 2
##   dataset   count
##   <chr>     <int>
## 1 AmeriFlux  3703
## 2 FLUXNET    6614
## 3 ICOS2018    336
## 4 ICOS2020   8362
```

Sites are linked to dataset

```
raw_df %>%
  group_by(SITE_ID, dataset) %>%
  summarise(count=n())
```

```
## `summarise()` has grouped output by 'SITE_ID'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 243 x 3
## # Groups:   SITE_ID [243]
##    SITE_ID dataset count
##    <chr>   <chr>   <int>
##  1 AR-SLu  FLUXNET    15
##  2 AR-Vir  FLUXNET    16
##  3 AT-Neu  FLUXNET   121
```

```
##  4 AU-Ade  FLUXNET    17
##  5 AU-ASM  FLUXNET    51
##  6 AU-Cpr  FLUXNET    48
##  7 AU-Cum  FLUXNET    25
##  8 AU-DaP  FLUXNET    55
##  9 AU-DaS  FLUXNET    74
## 10 AU-Dry  FLUXNET    45
## # ... with 233 more rows
```

Distribution of Land-cover Type by Site

SITE_IGBP(Land-cover Type): 11 SITE_ID(Sites): 243

```
raw_df %>%
  count(SITE_ID)
```

```
## # A tibble: 243 x 2
##    SITE_ID     n
##    <chr>   <int>
##  1 AR-SLu     15
##  2 AR-Vir     16
##  3 AT-Neu    121
##  4 AU-Ade     17
##  5 AU-ASM     51
##  6 AU-Cpr     48
##  7 AU-Cum     25
##  8 AU-DaP     55
##  9 AU-DaS     74
## 10 AU-Dry     45
## # ... with 233 more rows
```

```
raw_df
```

```
## # A tibble: 19,015 x 62
##    SITE_ID  year month TIMESTAMP dataset SITE_IGBP LOCATION_LAT LOCATION_LONG
##    <chr>   <dbl> <dbl>     <dbl> <chr>   <chr>            <dbl>         <dbl>
##  1 AR-SLu   2010     1    201001 FLUXNET MF               -33.5         -66.5
##  2 AR-SLu   2010     2    201002 FLUXNET MF               -33.5         -66.5
##  3 AR-SLu   2010     3    201003 FLUXNET MF               -33.5         -66.5
##  4 AR-SLu   2010     4    201004 FLUXNET MF               -33.5         -66.5
##  5 AR-SLu   2010     5    201005 FLUXNET MF               -33.5         -66.5
##  6 AR-SLu   2010     6    201006 FLUXNET MF               -33.5         -66.5
##  7 AR-SLu   2010     7    201007 FLUXNET MF               -33.5         -66.5
##  8 AR-SLu   2010     8    201008 FLUXNET MF               -33.5         -66.5
##  9 AR-SLu   2010     9    201009 FLUXNET MF               -33.5         -66.5
## 10 AR-SLu   2010    10    201010 FLUXNET MF               -33.5         -66.5
## # ... with 19,005 more rows, and 54 more variables: TA_F <dbl>, VPD_F <dbl>,
## #   P_F <dbl>, NETRAD <dbl>, NEE_VUT_REF <dbl>, NEE_VUT_REF_QC <dbl>,
## #   NEE_CUT_REF <dbl>, NEE_CUT_REF_QC <dbl>, GPP_NT_VUT_REF <dbl>,
## #   GPP_DT_VUT_REF <dbl>, GPP_NT_CUT_REF <dbl>, GPP_DT_CUT_REF <dbl>,
```

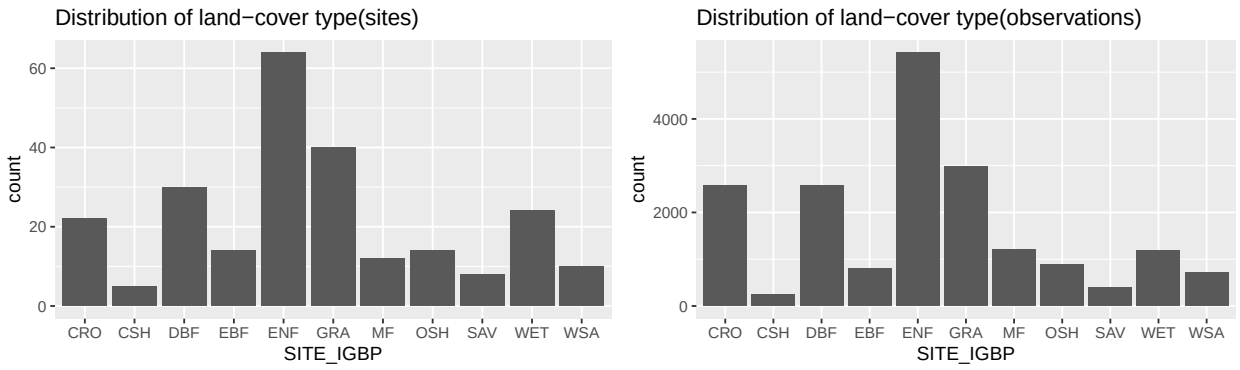
```
## #   RECO_NT_VUT_REF <dbl>, RECO_DT_VUT_REF <dbl>, RECO_NT_CUT_REF <dbl>,
## #   RECO_DT_CUT_REF <dbl>, time <chr>, ET <dbl>, `BESS-PAR` <dbl>,
## #   `BESS-PARdiff` <dbl>, `BESS-RSDN` <dbl>, `CSIF-SIFdaily` <dbl>, ...
```

```
library(dplyr)
site_igbp_distribution <- raw_df %>%
  dplyr::select(SITE_ID, SITE_IGBP) %>%
  group_by(SITE_IGBP) %>%
  summarise(count= n())

fig1 <- raw_df %>%
  dplyr::select(SITE_ID,SITE_IGBP) %>%
  unique() %>%
  group_by(SITE_IGBP) %>%
  summarise(count=n()) %>%
  ggplot(aes(x=SITE_IGBP, y=count)) +
  geom_bar(stat='identity') +
  labs(title = "Distribution of land-cover type(sites)")

fig2 <- site_igbp_distribution %>%
  ggplot(aes(x=SITE_IGBP, y=count)) +
  geom_bar(stat='identity') +
  labs(title = "Distribution of land-cover type(observations)")

grid.arrange(fig1, fig2, nrow = 1, ncol = 2)
```




```
site_igbp_distribution
```

```
## # A tibble: 11 x 2
##    SITE_IGBP count
##    <chr>     <int>
##  1 CRO        2574
##  2 CSH         252
##  3 DBF        2582
##  4 EBF         796
##  5 ENF        5422
##  6 GRA        2972
##  7 MF         1217
##  8 OSH         886
##  9 SAV         403
```

```
## 10 WET        1186
## 11 WSA         725
```

Add features that distinguish northern/southern hemisphere

```
raw_df_hemisphere <- raw_df %>%
  mutate(hemisphere = ifelse(LOCATION_LAT >= 0, "N", "S"))

raw_df_hemisphere
```

```
## # A tibble: 19,015 x 63
##    SITE_ID  year month TIMESTAMP dataset SITE_IGBP LOCATION_LAT LOCATION_LONG
##    <chr>   <dbl> <dbl>     <dbl> <chr>   <chr>            <dbl>         <dbl>
##  1 AR-SLu   2010     1    201001 FLUXNET MF               -33.5         -66.5
##  2 AR-SLu   2010     2    201002 FLUXNET MF               -33.5         -66.5
##  3 AR-SLu   2010     3    201003 FLUXNET MF               -33.5         -66.5
##  4 AR-SLu   2010     4    201004 FLUXNET MF               -33.5         -66.5
##  5 AR-SLu   2010     5    201005 FLUXNET MF               -33.5         -66.5
##  6 AR-SLu   2010     6    201006 FLUXNET MF               -33.5         -66.5
##  7 AR-SLu   2010     7    201007 FLUXNET MF               -33.5         -66.5
##  8 AR-SLu   2010     8    201008 FLUXNET MF               -33.5         -66.5
##  9 AR-SLu   2010     9    201009 FLUXNET MF               -33.5         -66.5
## 10 AR-SLu   2010    10    201010 FLUXNET MF               -33.5         -66.5
## # ... with 19,005 more rows, and 55 more variables: TA_F <dbl>, VPD_F <dbl>,
## #   P_F <dbl>, NETRAD <dbl>, NEE_VUT_REF <dbl>, NEE_VUT_REF_QC <dbl>,
## #   NEE_CUT_REF <dbl>, NEE_CUT_REF_QC <dbl>, GPP_NT_VUT_REF <dbl>,
## #   GPP_DT_VUT_REF <dbl>, GPP_NT_CUT_REF <dbl>, GPP_DT_CUT_REF <dbl>,
## #   RECO_NT_VUT_REF <dbl>, RECO_DT_VUT_REF <dbl>, RECO_NT_CUT_REF <dbl>,
## #   RECO_DT_CUT_REF <dbl>, time <chr>, ET <dbl>, `BESS-PAR` <dbl>,
## #   `BESS-PARdiff` <dbl>, `BESS-RSDN` <dbl>, `CSIF-SIFdaily` <dbl>, ...
```

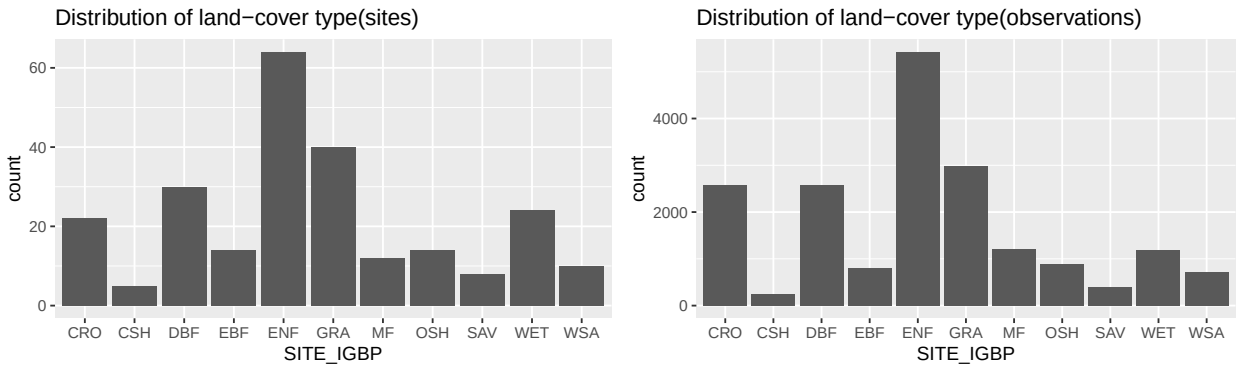
```
library(dplyr)
site_igbp_distribution <- raw_df_hemisphere %>%

  dplyr::select(SITE_ID, SITE_IGBP) %>%
  group_by(SITE_IGBP) %>%
  summarise(count= n())

fig1 <- raw_df %>%
  dplyr::select(SITE_ID,SITE_IGBP) %>%
  unique() %>%
  group_by(SITE_IGBP) %>%
  summarise(count=n()) %>%
  ggplot(aes(x=SITE_IGBP, y=count)) +
  geom_bar(stat='identity') +
  labs(title = "Distribution of land-cover type(sites)")

fig2 <- site_igbp_distribution %>%
  ggplot(aes(x=SITE_IGBP, y=count)) +
  geom_bar(stat='identity') +
  labs(title = "Distribution of land-cover type(observations)")
```

```
grid.arrange(fig1, fig2, nrow = 1, ncol = 2)
```



Observe categorical variables

Since they only have 243 rows by removing all the duplicates, all the categorical has unique values in each site.

```
raw_df_categorical <- raw_df_hemisphere %>%
  dplyr::select(SITE_ID, SITE_IGBP, dataset, MODIS_IGBP, MODIS_PFT,
         koppen_sub, koppen, hemisphere) %>%
  distinct() #drop duplicates

head(raw_df_categorical)
```

```
## # A tibble: 6 x 8
##   SITE_ID SITE_IGBP dataset MODIS_IGBP MODIS_PFT koppen_sub koppen    hemisphere
##   <chr>   <chr>     <chr>   <chr>      <chr>     <chr>      <chr>     <chr>
## 1 AR-SLu  MF        FLUXNET OSH        SH        BSk        Arid      S
## 2 AR-Vir  ENF       FLUXNET SAV        SA        Cfa        Temperate S
## 3 AT-Neu  GRA       FLUXNET MF         MF        Dfb        Cold      N
## 4 AU-ASM  SAV       FLUXNET CSH        SH        BWh        Arid      S
## 5 AU-Ade  WSA       FLUXNET GRA        GRA       Aw         Tropical  S
## 6 AU-Cpr  SAV       FLUXNET CSH        SH        BWk        Arid      S
```

```
raw_df_categorical %>%
  distinct(koppen) %>%
  as.list()
```

```
## $koppen
## [1] "Arid"      "Temperate" "Cold"      "Tropical"  "Polar"
```

Unique values of categorical variables

- SITE_ID: 243
- SITE_IGBP: 11
  - MF, ENF, GRA, SAV, WSA, EBF, WET, OSH, DBF, CRO, CSH

- MODIS_IGBP: 14
  - OSH, SAV, MF, CSH, GRA, ENF, EBF, WSA, DBF, CRO, WAT, BSV, URB, WET
- MODIS_PFT: 14
  - SH, SA, MF, GRA, ENF, EBF, DBF, CRO, Other
- koppen_sub: 21
  - BSk, Cfa, Dfb, BWh, Aw, BWk, BSh, Csa, Cfb, Am, Dfc, Dwb, Dwa, Cwa, Dwc, ET, Dsb, Af, Dsc, Csb, Dfa
- koppen: 5
  - Arid, Temperate, Cold, Tropical, Polar

Convert to monthly average dataset

Create a new df of monthly average across sites(SITE_ID)

row: 2781

```
SITE_month_df <- raw_df_hemisphere %>%
  group_by(SITE_ID, SITE_IGBP, month) %>%
  summarise(TA_F_avg = mean(TA_F, na.rm = T),VPD_F_avg = mean(VPD_F, na.rm = T),
            P_F_avg = mean(P_F, na.rm = T),NETRAD_avg = mean(NETRAD, na.rm = T),
            NEE_VUT_REF_avg = mean(NEE_VUT_REF, na.rm = T),
            NEE_VUT_REF_QC_avg = mean(NEE_VUT_REF_QC, na.rm = T),
            NEE_CUT_REF_avg = mean(NEE_CUT_REF, na.rm = T),
            NEE_CUT_REF_QC_avg = mean(NEE_CUT_REF_QC, na.rm = T),
            GPP_NT_VUT_REF_avg = mean(GPP_NT_VUT_REF, na.rm = T),
            GPP_DT_VUT_REF_avg = mean(GPP_DT_VUT_REF, na.rm = T),
            GPP_NT_CUT_REF_avg = mean(GPP_NT_CUT_REF, na.rm = T),
            GPP_DT_CUT_REF_avg = mean(GPP_DT_CUT_REF, na.rm = T),
            RECO_NT_VUT_REF_avg = mean(RECO_NT_VUT_REF, na.rm = T),
            RECO_DT_VUT_REF_avg = mean(RECO_DT_VUT_REF, na.rm = T),
            RECO_NT_CUT_REF_avg = mean(RECO_NT_CUT_REF, na.rm = T),
            RECO_DT_CUT_REF_avg = mean(RECO_DT_CUT_REF, na.rm = T),
            ET_avg = mean(ET, na.rm = T),
            `BESS-PAR_avg` = mean(`BESS-PAR`, na.rm = T),
            `BESS-PARdiff_avg` = mean(`BESS-PARdiff`, na.rm = T),
            `BESS-RSDN_avg` = mean(`BESS-RSDN`, na.rm = T),
            `CSIF-SIFdaily_avg` = mean(`CSIF-SIFdaily`, na.rm = T),
            `CSIF-SIFinst_avg` = mean(`CSIF-SIFinst`, na.rm = T),
            PET_avg = mean(PET, na.rm = T),Ts_avg = mean(Ts, na.rm = T),
            Tmean_avg = mean(Tmean, na.rm = T),
            prcp_avg = mean(prcp, na.rm = T),
            vpd_avg = mean(vpd, na.rm = T),
            `prcp-lag3_avg` = mean(`prcp-lag3`, na.rm = T),
            `ESACCI-sm_avg` = mean(`ESACCI-sm`, na.rm = T),
            b1_avg = mean(b1, na.rm = T),b2_avg = mean(b2, na.rm = T),
            b3_avg = mean(b3, na.rm = T),b4_avg = mean(b4, na.rm = T),
            b5_avg = mean(b5, na.rm = T),b6_avg = mean(b6, na.rm = T),
            b7_avg = mean(b7, na.rm = T),EVI_avg = mean(EVI, na.rm = T),
            GCI_avg = mean(GCI, na.rm = T),NDVI_avg = mean(NDVI, na.rm = T),
            NDWI_avg = mean(NDWI, na.rm = T),NIRv_avg = mean(NIRv, na.rm = T),
```

```
            kNDVI_avg = mean(kNDVI, na.rm = T),
            Percent_Snow_avg = mean(Percent_Snow, na.rm = T),
            Fpar_avg = mean(Fpar, na.rm = T),Lai_avg = mean(Lai, na.rm = T),
            LST_Day_avg = mean(LST_Day, na.rm = T),
            LST_Night_avg = mean(LST_Night, na.rm = T),
            CO2_concentration_avg = mean(CO2_concentration, na.rm = T)
                        )
```

```
## `summarise()` has grouped output by 'SITE_ID', 'SITE_IGBP'. You can override
## using the `.groups` argument.
```

```
# Merge categorical variables

raw_df_categorical_ <- raw_df_categorical %>%
  dplyr::select(-c(SITE_IGBP)) # drop SITE_IGBP to avoid two variables in one df
SITE_month_df <- merge(x = SITE_month_df, y = raw_df_categorical_,
                       by="SITE_ID", all.x = TRUE)

# Add longitude/latitude to monthly df
raw_df_geo <- raw_df_hemisphere %>% dplyr::select(SITE_ID, LOCATION_LAT, LOCATION_LONG)
SITE_month_df <- merge(x = SITE_month_df, y = raw_df_geo,
                       by="SITE_ID", all.x = TRUE)

# raw_df_ts_month <- raw_df %>% as_tsibble(index=month)
```

Create a new df of monthly average across land-cover types(SITE_IGBP)

Categorical variables are excluded due to its dependency to each site.

```
IGBP_month_df <- raw_df_hemisphere %>%
  group_by(SITE_IGBP, hemisphere, month) %>%
  summarise(TA_F_avg = mean(TA_F, na.rm = T),VPD_F_avg = mean(VPD_F, na.rm = T),
            P_F_avg = mean(P_F, na.rm = T),NETRAD_avg = mean(NETRAD, na.rm = T),
            NEE_VUT_REF_avg = mean(NEE_VUT_REF, na.rm = T),
            NEE_VUT_REF_QC_avg = mean(NEE_VUT_REF_QC, na.rm = T),
            NEE_CUT_REF_avg = mean(NEE_CUT_REF, na.rm = T),
            NEE_CUT_REF_QC_avg = mean(NEE_CUT_REF_QC, na.rm = T),
            GPP_NT_VUT_REF_avg = mean(GPP_NT_VUT_REF, na.rm = T),
            GPP_DT_VUT_REF_avg = mean(GPP_DT_VUT_REF, na.rm = T),
            GPP_NT_CUT_REF_avg = mean(GPP_NT_CUT_REF, na.rm = T),
            GPP_DT_CUT_REF_avg = mean(GPP_DT_CUT_REF, na.rm = T),
            RECO_NT_VUT_REF_avg = mean(RECO_NT_VUT_REF, na.rm = T),
            RECO_DT_VUT_REF_avg = mean(RECO_DT_VUT_REF, na.rm = T),
            RECO_NT_CUT_REF_avg = mean(RECO_NT_CUT_REF, na.rm = T),
            RECO_DT_CUT_REF_avg = mean(RECO_DT_CUT_REF, na.rm = T),
            ET_avg = mean(ET, na.rm = T),
            `BESS-PAR_avg` = mean(`BESS-PAR`, na.rm = T),
            `BESS-PARdiff_avg` = mean(`BESS-PARdiff`, na.rm = T),
            `BESS-RSDN_avg` = mean(`BESS-RSDN`, na.rm = T),
            `CSIF-SIFdaily_avg` = mean(`CSIF-SIFdaily`, na.rm = T),
            `CSIF-SIFinst_avg` = mean(`CSIF-SIFinst`, na.rm = T),
            PET_avg = mean(PET, na.rm = T),Ts_avg = mean(Ts, na.rm = T),
```

```
            Tmean_avg = mean(Tmean, na.rm = T),
            prcp_avg = mean(prcp, na.rm = T),
            vpd_avg = mean(vpd, na.rm = T),
            `prcp-lag3_avg` = mean(`prcp-lag3`, na.rm = T),
            `ESACCI-sm_avg` = mean(`ESACCI-sm`, na.rm = T),
            b1_avg = mean(b1, na.rm = T),b2_avg = mean(b2, na.rm = T),
            b3_avg = mean(b3, na.rm = T),b4_avg = mean(b4, na.rm = T),
            b5_avg = mean(b5, na.rm = T),b6_avg = mean(b6, na.rm = T),
            b7_avg = mean(b7, na.rm = T),EVI_avg = mean(EVI, na.rm = T),
            GCI_avg = mean(GCI, na.rm = T),NDVI_avg = mean(NDVI, na.rm = T),
            NDWI_avg = mean(NDWI, na.rm = T),NIRv_avg = mean(NIRv, na.rm = T),
            kNDVI_avg = mean(kNDVI, na.rm = T),
            Percent_Snow_avg = mean(Percent_Snow, na.rm = T),
            Fpar_avg = mean(Fpar, na.rm = T),Lai_avg = mean(Lai, na.rm = T),
            LST_Day_avg = mean(LST_Day, na.rm = T),
            LST_Night_avg = mean(LST_Night, na.rm = T),
            CO2_concentration_avg = mean(CO2_concentration, na.rm = T)
                        )
```

```
## `summarise()` has grouped output by 'SITE_IGBP', 'hemisphere'. You can override
## using the `.groups` argument.
```

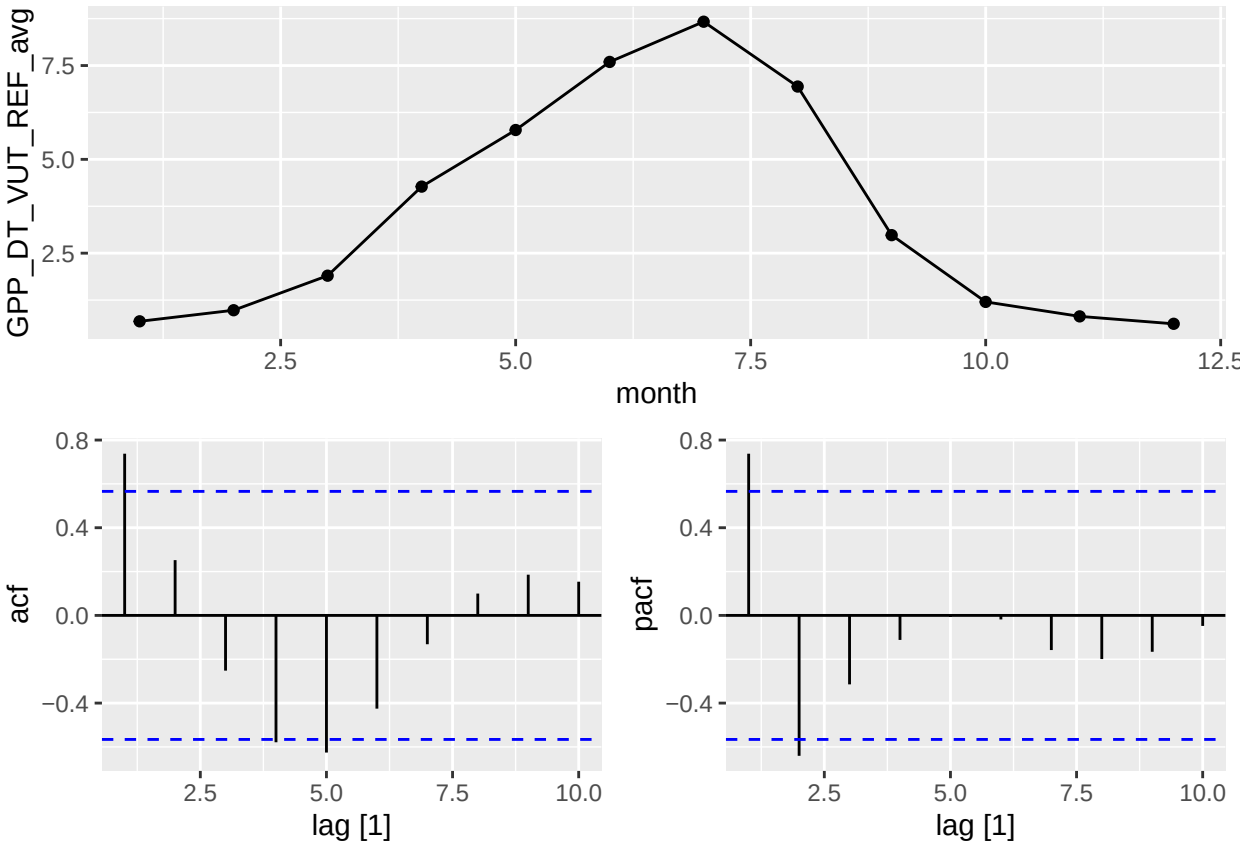
```
# raw_df_ts_month <- raw_df %>% as_tsibble(index=month)
```

Export df to csv

```
# write.csv(IGBP_month_df,
#           "../../../data/datasets/SITE_IGBP_month_df.csv",row.names=FALSE)
# write.csv(SITE_month_df,
#           "../../../data/datasets/SITE_month_df.csv", row.names=FALSE)
```

Monthly transition of GPP and NEE by land-cover type

```
IGBP_month_df %>% subset(SITE_IGBP == "CRO") %>%
  subset(hemisphere == "N") %>%
  as_tsibble(index = month) %>%
  gg_tsdisplay(GPP_DT_VUT_REF_avg, plot_type="partial")
```

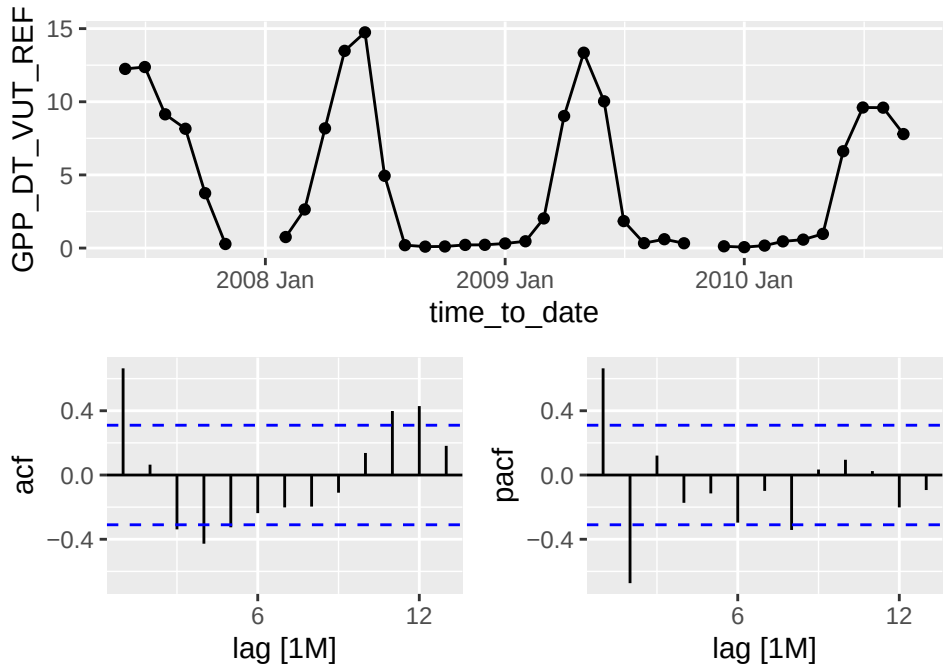



```
# GPP_DT_VUT_REF in DE-Seh
print("GPP_DT_VUT_REF in DE-Seh")
```

```
## [1] "GPP_DT_VUT_REF in DE-Seh"
```

```
raw_df_hemisphere %>% subset(SITE_ID == "DE-Seh") %>%
  subset(hemisphere == "N") %>%
  mutate(time_to_date = yearmonth(as.Date(time, format = "%m/%d/%y"))) %>%
  as_tsibble(index = time_to_date) %>%
  tsibble::fill_gaps() %>%
  gg_tsdisplay(GPP_DT_VUT_REF, plot_type="partial")
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
```

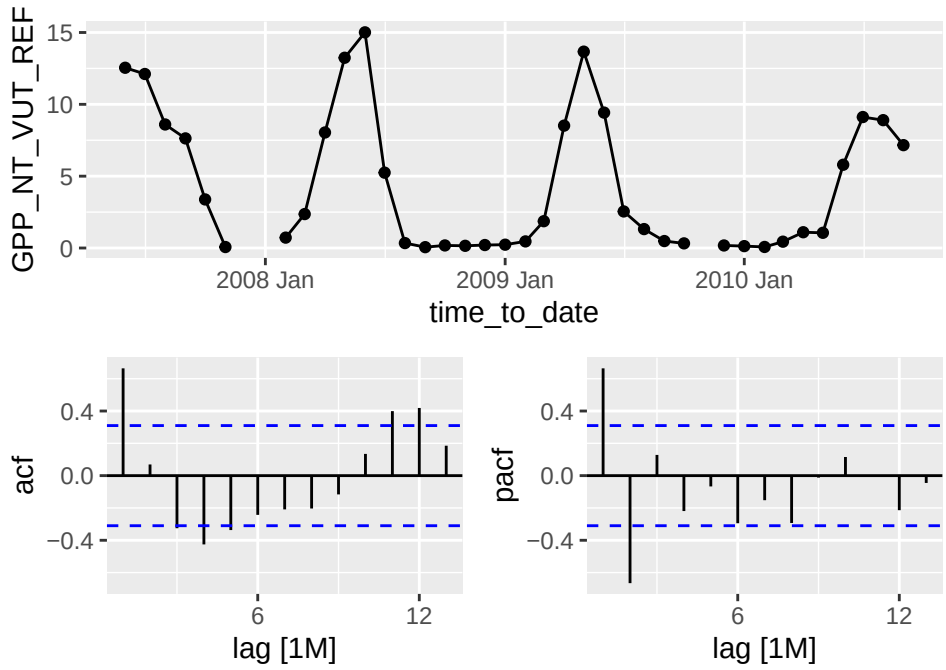



```
# GPP_NT_VUT_REF in DE-Seh
print("GPP_NT_VUT_REF in DE-Seh")
```

```
## [1] "GPP_NT_VUT_REF in DE-Seh"
```

```
raw_df_hemisphere %>% subset(SITE_ID == "DE-Seh") %>%
  subset(hemisphere == "N") %>%
  mutate(time_to_date = yearmonth(as.Date(time, format = "%m/%d/%y"))) %>%
  as_tsibble(index = time_to_date) %>%
  tsibble::fill_gaps() %>%
  gg_tsdisplay(GPP_NT_VUT_REF, plot_type="partial")
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
```

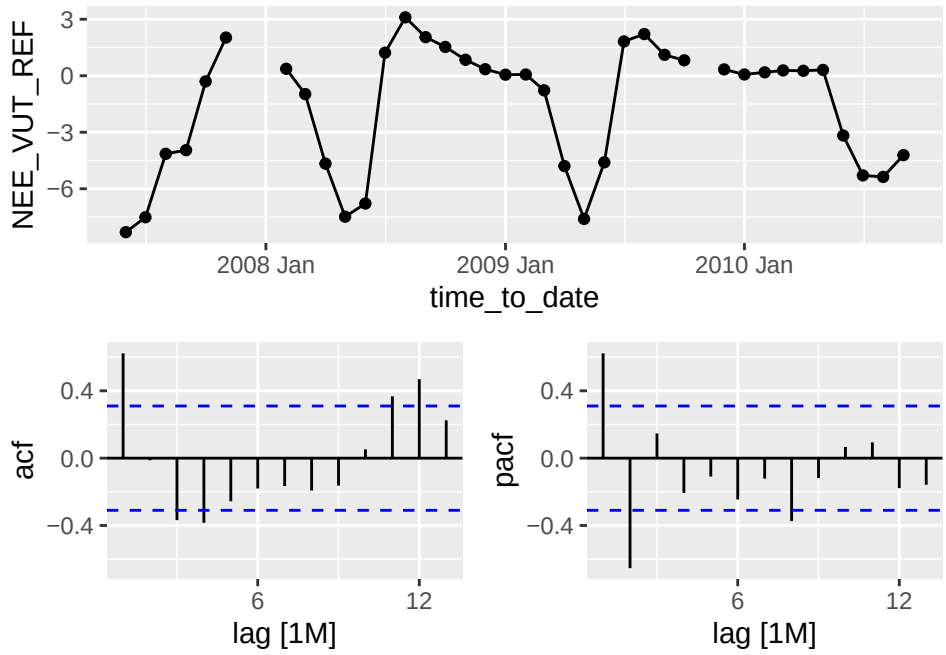



```
# NEE_VUT_REF in DE-Seh
print("NEE_VUT_REF in DE-Seh")
```

```
## [1] "NEE_VUT_REF in DE-Seh"
```

```
raw_df_hemisphere %>% subset(SITE_ID == "DE-Seh") %>%
  subset(hemisphere == "N") %>%
  mutate(time_to_date = yearmonth(as.Date(time, format = "%m/%d/%y"))) %>%
  as_tsibble(index = time_to_date) %>%
  tsibble::fill_gaps() %>%
  gg_tsdisplay(NEE_VUT_REF, plot_type="partial")
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
```

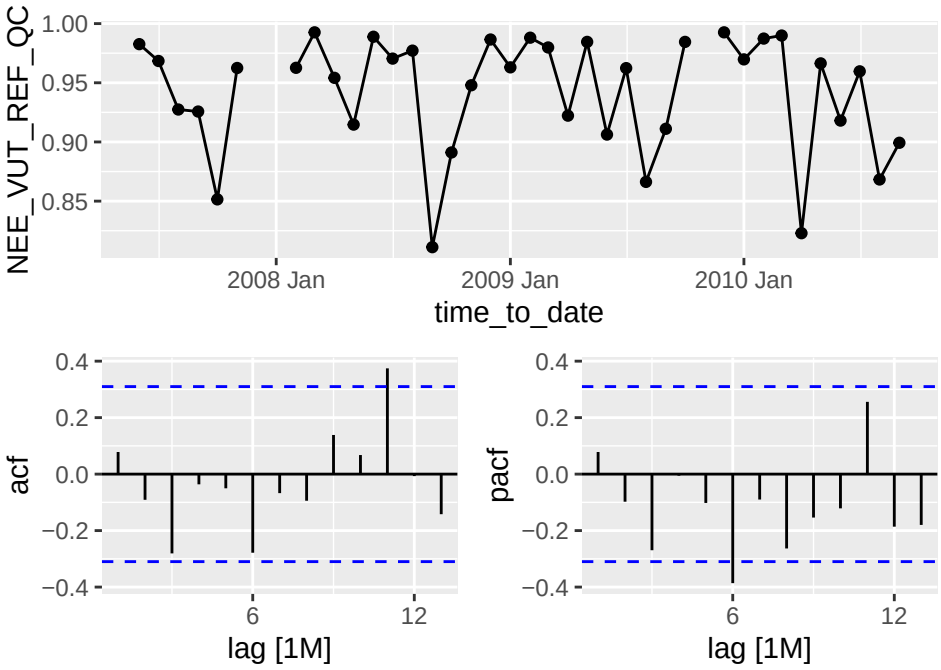



```
# NEE_VUT_REF_QC in DE-Seh
print("NEE_VUT_REF_QC in DE-Seh")
```

```
## [1] "NEE_VUT_REF_QC in DE-Seh"
```

```
raw_df_hemisphere %>% subset(SITE_ID == "DE-Seh") %>%
  subset(hemisphere == "N") %>%
  mutate(time_to_date = yearmonth(as.Date(time, format = "%m/%d/%y"))) %>%
  as_tsibble(index = time_to_date) %>%
  tsibble::fill_gaps() %>%
  gg_tsdisplay(NEE_VUT_REF_QC, plot_type="partial")
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
```



Appendix

Summary statistics of features

Rows: 19015 Columns: 62

```
describe(raw_df)
```

```
## raw_df
##
##  62  Variables      19015  Observations
## ---------------------------------------------------------------------------
## SITE_ID
##        n  missing distinct
##    19015        0      243
##
## lowest : AR-SLu AR-Vir AT-Neu AU-Ade AU-ASM, highest: US-Wi4 US-Wjs US-Wkg US-WPT ZM-Mon
## ---------------------------------------------------------------------------
## year
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0       20    0.997     2011    5.927     2002     2004
##      .25      .50      .75      .90      .95
##     2006     2011     2015     2018     2019
##
## lowest : 2001 2002 2003 2004 2005, highest: 2016 2017 2018 2019 2020
##
## Value       2001  2002  2003  2004  2005  2006  2007  2008  2009  2010  2011
## Frequency    444   619   730   927  1038  1017  1092  1179  1118  1147  1191
## Proportion 0.023 0.033 0.038 0.049 0.055 0.053 0.057 0.062 0.059 0.060 0.063
```

```
##
## Value       2012  2013  2014  2015  2016  2017  2018  2019  2020
## Frequency   1275  1212  1188   876   936   897   720   705   704
## Proportion 0.067 0.064 0.062 0.046 0.049 0.047 0.038 0.037 0.037
## ---------------------------------------------------------------------------
## month
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0       12    0.993    6.591    3.898        1        2
##      .25      .50      .75      .90      .95
##        4        7       10       11       12
##
## lowest :  1  2  3  4  5, highest:  8  9 10 11 12
##
## Value          1     2     3     4     5     6     7     8     9    10    11
## Frequency   1422  1479  1545  1560  1633  1629  1631  1682  1668  1673  1600
## Proportion 0.075 0.078 0.081 0.082 0.086 0.086 0.086 0.088 0.088 0.088 0.084
##
## Value         12
## Frequency   1493
## Proportion 0.079
## ---------------------------------------------------------------------------
## TIMESTAMP
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0      240        1   201069    592.8   200210   200402
##      .25      .50      .75      .90      .95
##   200612   201103   201502   201804   201908
##
## lowest : 200101 200102 200103 200104 200105, highest: 202008 202009 202010 202011 202012
## ---------------------------------------------------------------------------
## dataset
##        n  missing distinct
##    19015        0        4
##
## Value      AmeriFlux   FLUXNET  ICOS2018  ICOS2020
## Frequency       3703      6614       336      8362
## Proportion     0.195     0.348     0.018     0.440
## ---------------------------------------------------------------------------
## SITE_IGBP
##        n  missing distinct
##    19015        0       11
##
## lowest : CRO CSH DBF EBF ENF, highest: MF  OSH SAV WET WSA
##
## Value        CRO   CSH   DBF   EBF   ENF   GRA    MF   OSH   SAV   WET   WSA
## Frequency   2574   252  2582   796  5422  2972  1217   886   403  1186   725
## Proportion 0.135 0.013 0.136 0.042 0.285 0.156 0.064 0.047 0.021 0.062 0.038
## ---------------------------------------------------------------------------
## LOCATION_LAT
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0      243        1    42.19    16.44   -12.49    31.79
##      .25      .50      .75      .90      .95
##    39.94    45.95    50.89    56.10    64.17
##
## lowest : -37.42590 -37.42220 -36.67320 -36.64990 -35.65660
```

```
## highest:  70.46960  70.82914  74.47328  74.48143  78.18600
## ---------------------------------------------------------------------------
## LOCATION_LONG
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0      242        1   -21.17    73.53 -120.966 -109.942
##      .25      .50      .75      .90      .95
##  -90.080    5.744   13.513   29.610  131.152
##
## lowest : -157.4089 -155.7503 -147.8555 -147.4876 -125.3336
## highest:  147.4943  148.1517  148.4746  150.7236  161.3414
## ---------------------------------------------------------------------------
## TA_F
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0    14594        1    10.34    10.72   -5.710   -1.971
##      .25      .50      .75      .90      .95
##    3.874   10.966   17.159   22.520   25.248
##
## lowest : -31.455 -28.628 -28.290 -28.173 -27.865
## highest:  32.618  32.707  32.847  32.992  33.400
## ---------------------------------------------------------------------------
## VPD_F
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0     9775        1    5.432    5.098    0.495    0.832
##      .25      .50      .75      .90      .95
##    1.884    4.008    6.998   11.903   16.286
##
## lowest :  0.007  0.022  0.034  0.046  0.060, highest: 37.228 37.798 38.316 39.361 45.384
## ---------------------------------------------------------------------------
## P_F
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18662      353     5356        1    2.095    1.982    0.025    0.175
##      .25      .50      .75      .90      .95
##    0.727    1.619    2.871    4.424    5.762
##
## lowest :  0.000  0.001  0.002  0.003  0.004, highest: 20.696 22.810 23.365 24.635 25.172
## ---------------------------------------------------------------------------
## NETRAD
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    16322     2693    16321        1     81.5    69.64   -7.053    1.438
##      .25      .50      .75      .90      .95
##   25.833   83.546  131.353  160.143  175.072
##
## lowest : -67.89401 -62.49031 -59.65575 -57.19109 -52.67848
## highest: 287.13336 296.88608 301.63435 307.46765 307.99158
## ---------------------------------------------------------------------------
## NEE_VUT_REF
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0    18934        1  -0.6366    2.187 -4.91723 -3.42104
##      .25      .50      .75      .90      .95
## -1.49466 -0.03468  0.63834  1.31770  1.80338
##
## lowest : -14.02970 -13.83060 -13.71760 -13.37000 -13.29610
## highest:   7.01448   7.44355   8.05616   8.71916   8.81725
## ---------------------------------------------------------------------------
```

```
## NEE_VUT_REF_QC
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0     1002        1   0.9561  0.04623   0.8548   0.8905
##      .25      .50      .75      .90      .95
##   0.9395   0.9718   0.9886   0.9965   0.9987
##
## lowest : 0.431452 0.461022 0.501344 0.556452 0.614247
## highest: 0.999256 0.999282 0.999306 0.999328 1.000000
## ------------------------------------------------------------------------------
## NEE_CUT_REF
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0    18907        1  -0.6406    2.185 -4.90614 -3.42255
##      .25      .50      .75      .90      .95
## -1.48760 -0.03645  0.63756  1.31050  1.78472
##
## lowest : -14.08820 -13.90990 -13.83730 -13.43980 -13.34270
## highest:   7.05364   7.41292   7.49866   7.82519   8.75457
## ------------------------------------------------------------------------------
## NEE_CUT_REF_QC
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0      883        1   0.9576  0.04409   0.8594   0.8931
##      .25      .50      .75      .90      .95
##   0.9415   0.9724   0.9886   0.9965   0.9987
##
## lowest : 0.800287 0.800403 0.800694 0.801075 0.801339
## highest: 0.999256 0.999282 0.999306 0.999328 1.000000
## ------------------------------------------------------------------------------
## GPP_NT_VUT_REF
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0    18855        1    3.634     4.11 -0.06870  0.01167
##      .25      .50      .75      .90      .95
##  0.40645  2.28575  6.06953  9.34163 11.16933
##
## lowest : -0.999454 -0.951902 -0.941732 -0.939231 -0.928428
## highest: 23.035600 23.229300 23.253600 23.286900 23.775600
## ------------------------------------------------------------------------------
## GPP_DT_VUT_REF
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0    18583        1    3.516    3.896  0.01442  0.06147
##      .25      .50      .75      .90      .95
##  0.43092  2.24238  5.82667  9.00748 10.62423
##
## lowest :  0.00000000  0.00000467  0.00000486  0.00001580  0.00001890
## highest: 23.29540000 23.39650000 23.41760000 23.79160000 24.06010000
## ------------------------------------------------------------------------------
## GPP_NT_CUT_REF
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0    18860        1    3.627    4.103 -0.06668  0.01335
##      .25      .50      .75      .90      .95
##  0.40735  2.27758  6.06802  9.31733 11.14000
##
## lowest : -1.223370 -1.059510 -1.044810 -1.032150 -0.955025
## highest: 22.793300 23.322800 23.347500 23.525300 23.888800
## ------------------------------------------------------------------------------
```

```
## GPP_DT_CUT_REF
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0    18534        1    3.512    3.884  0.01458  0.06359
##      .25      .50      .75      .90      .95
##  0.43464  2.25236  5.81980  8.97261 10.60416
##
## lowest :  0.00000000  0.00000899  0.00000974  0.00001060  0.00001940
## highest: 23.02800000 23.37300000 23.73030000 24.07310000 24.78880000
## ---------------------------------------------------------------------------
## RECO_NT_VUT_REF
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0    18770        1    2.986    2.749   0.1946   0.4017
##      .25      .50      .75      .90      .95
##   0.9853   2.2191   4.3961   6.5565   8.1156
##
## lowest : 0.00000e+00 3.49000e-09 4.47000e-09 3.42000e-08 4.04000e-08
## highest: 1.86266e+01 1.86878e+01 1.91984e+01 1.95403e+01 1.96794e+01
## ---------------------------------------------------------------------------
## RECO_DT_VUT_REF
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19014        1    18816        1    2.931    2.611   0.2957   0.4771
##      .25      .50      .75      .90      .95
##   1.0106   2.2123   4.3329   6.3974   7.7122
##
## lowest : 1.94896e-04 4.33736e-03 4.45924e-03 4.84733e-03 5.38437e-03
## highest: 1.51963e+01 1.53055e+01 1.62103e+01 1.73310e+01 1.77965e+01
## ---------------------------------------------------------------------------
## RECO_NT_CUT_REF
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0    18794        1    2.965    2.735   0.1890   0.3946
##      .25      .50      .75      .90      .95
##   0.9733   2.2060   4.3738   6.5235   8.1121
##
## lowest : 0.00000e+00 3.49000e-09 4.47000e-09 3.49000e-08 4.04000e-08
## highest: 1.87690e+01 1.91544e+01 1.95403e+01 1.95475e+01 1.97785e+01
## ---------------------------------------------------------------------------
## RECO_DT_CUT_REF
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19014        1    18813        1    2.916      2.6   0.2936   0.4734
##      .25      .50      .75      .90      .95
##   1.0060   2.1998   4.3056   6.3702   7.6631
##
## lowest :  0.00433736  0.00445924  0.00511637  0.00725787  0.00790458
## highest: 15.33190000 15.48180000 16.39560000 17.53900000 17.79650000
## ---------------------------------------------------------------------------
## time
##        n  missing distinct
##    19015        0      240
##
## lowest : 1/31/01 1/31/02 1/31/03 1/31/04 1/31/05
## highest: 9/30/16 9/30/17 9/30/18 9/30/19 9/30/20
## ---------------------------------------------------------------------------
## ET
##        n  missing distinct     Info     Mean      Gmd      .05      .10
```

```
##    18983       32    18337        1    3.666    3.227   0.1080   0.2531
##      .25      .50      .75      .90      .95
##   1.2016   3.1427   5.6596   7.6979   8.9099
##
## lowest :  0.0000000  0.0000105  0.0000108  0.0000121  0.0000148
## highest: 14.3170140 14.5098940 14.6386310 14.7813740 14.8986400
## ---------------------------------------------------------------------------
## BESS-PAR
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0      415        1    79.33    56.64       12       18
##      .25      .50      .75      .90      .95
##       39       77      106      129      145
##
## lowest :   0   1   2   3   4, highest: 588 597 599 600 611
## ---------------------------------------------------------------------------
## BESS-PARdiff
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0      216        1    33.85    22.54        7       10
##      .25      .50      .75      .90      .95
##       18       32       45       52       56
##
## lowest :   0   1   2   3   4, highest: 226 232 235 240 241
## ---------------------------------------------------------------------------
## BESS-RSDN
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0      374        1    157.8    100.8       25       37
##      .25      .50      .75      .90      .95
##       81      163      225      271      301
##
## lowest :   0   1   2   3   4, highest: 369 370 371 372 373
## ---------------------------------------------------------------------------
## CSIF-SIFdaily
##         n   missing  distinct      Info      Mean       Gmd       .05       .10
##     18894       121     17919         1    0.1517    0.1517 -0.000505  0.010790
##       .25       .50       .75       .90       .95
##  0.041752  0.102851  0.235123  0.370362  0.440347
##
## lowest : -0.02919996 -0.02814803 -0.02656131 -0.02655265 -0.02044134
## highest:  0.68423855  0.68502396  0.68649540  0.69930070  0.70735025
## ---------------------------------------------------------------------------
## CSIF-SIFinst
##         n   missing  distinct      Info      Mean       Gmd       .05       .10
##     18894       121     17772         1    0.4234    0.3902 -0.002229  0.036781
##       .25       .50       .75       .90       .95
##  0.148847  0.322552  0.641604  0.962354  1.140990
##
## lowest : -0.11166026 -0.10104087 -0.09271585 -0.08875447 -0.08076532
## highest:  1.72390760  1.73143880  1.75219540  1.75410040  1.79071280
## ---------------------------------------------------------------------------
## PET
##          n    missing   distinct       Info       Mean        Gmd        .05
##      18914        101      17748          1  -0.007306   0.006741 -0.0222709
##        .10        .25        .50        .75        .90        .95
## -0.0164077 -0.0091421 -0.0055377 -0.0026242 -0.0011704 -0.0007637
```

```
##
## lowest : -0.05542774 -0.05510548 -0.05494109 -0.05482972 -0.05447997
## highest:  0.00003210  0.00003400  0.00003460  0.00003720  0.00004160
## ---------------------------------------------------------------------------
## Ts
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18914      101    17589        1    283.6    12.11    265.3    269.6
##      .25      .50      .75      .90      .95
##    276.3    284.2    291.0    297.7    300.6
##
## lowest : 239.4686 239.6578 242.1032 242.7621 243.2340
## highest: 309.3234 309.5114 309.7074 310.0883 310.2855
## ---------------------------------------------------------------------------
## Tmean
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18914      101    17565        1    283.6    11.13    266.4    270.7
##      .25      .50      .75      .90      .95
##    277.0    284.3    290.7    296.2    298.9
##
## lowest : 241.8833 244.6789 245.3720 245.4935 245.5205
## highest: 305.7779 305.7994 305.8423 306.1157 306.8248
## ---------------------------------------------------------------------------
## prcp
##         n   missing  distinct      Info      Mean       Gmd       .05       .10
##     18914       101     17622         1  0.002451  0.002055 0.0001412 0.0004147
##       .25       .50       .75       .90       .95
## 0.0010851 0.0020328 0.0032729 0.0047820 0.0060673
##
## lowest : 0.000000540 0.000000555 0.000000586 0.000000608 0.000000623
## highest: 0.022791667 0.022880048 0.023600347 0.023827540 0.023946350
## ---------------------------------------------------------------------------
## vpd
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18914      101    17821        1   0.5254   0.5012  0.06597  0.09283
##      .25      .50      .75      .90      .95
##  0.17472  0.37054  0.65985  1.18798  1.64899
##
## lowest : 0.01310228 0.01468146 0.01750002 0.01795330 0.01799651
## highest: 3.59606310 3.80632110 3.86339000 3.94436570 4.21067430
## ---------------------------------------------------------------------------
## prcp-lag3
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18914      101    17805        1 0.007331 0.005093 0.001072 0.002054
##      .25      .50      .75      .90      .95
## 0.004168 0.006535 0.009392 0.012903 0.015871
##
## lowest : 0.00000211 0.00000229 0.00000244 0.00000257 0.00000261
## highest: 0.05334827 0.05434239 0.05618351 0.06139616 0.06408255
## ---------------------------------------------------------------------------
## ESACCI-sm
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    17405     1610    13985        1   0.2566  0.07281   0.1339   0.1658
##      .25      .50      .75      .90      .95
##   0.2168   0.2650   0.3011   0.3332   0.3529
```

```
##
## lowest : 0.00000838 0.04904429 0.05008446 0.05073670 0.05515835
## highest: 0.41661453 0.41684616 0.41879398 0.41975582 0.42007230
## ---------------------------------------------------------------------------
## MODIS_LC
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0       14     0.98     7.97    4.218        1        1
##      .25      .50      .75      .90      .95
##        5        9       11       12       12
##
## lowest :  1  2  4  5  6, highest: 11 12 13 16 17
##
## Value          1     2     4     5     6     7     8     9    10    11    12
## Frequency   1973   306  1576  2032    99   894  2612  2475  2145   173  4197
## Proportion 0.104 0.016 0.083 0.107 0.005 0.047 0.137 0.130 0.113 0.009 0.221
##
## Value         13    16    17
## Frequency     73     8   452
## Proportion 0.004 0.000 0.024
## ---------------------------------------------------------------------------
## b1
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18609      406    16720        1  0.08678  0.06897  0.02623  0.03014
##      .25      .50      .75      .90      .95
##  0.04208  0.06154  0.09900  0.17116  0.23396
##
## lowest : 0.01132731 0.01195806 0.01283010 0.01378667 0.01388188
## highest: 0.87467510 0.88530600 0.88612750 0.88924116 0.92057824
## ---------------------------------------------------------------------------
## b2
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18616      399    16698        1   0.2466  0.08636   0.1413   0.1625
##      .25      .50      .75      .90      .95
##   0.1933   0.2371   0.2930   0.3506   0.3789
##
## lowest : 0.00000000 0.04759337 0.06271509 0.06848542 0.07399469
## highest: 0.80770830 0.81104064 0.81478095 0.82439053 0.87086660
## ---------------------------------------------------------------------------
## b3
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18609      406    16481        1  0.05725  0.05532  0.01485  0.01771
##      .25      .50      .75      .90      .95
##  0.02400  0.03386  0.05265  0.09979  0.21867
##
## lowest : 0.005035000 0.005451613 0.005491508 0.005632258 0.005638444
## highest: 0.895115100 0.900761300 0.905036800 0.907366600 0.912842900
## ---------------------------------------------------------------------------
## b4
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18609      406    16622        1  0.08417   0.0584  0.03395  0.03839
##      .25      .50      .75      .90      .95
##  0.04768  0.06319  0.08669  0.14148  0.22247
##
## lowest : 0.01439973 0.01603387 0.01698667 0.01717167 0.01735941
```

```
## highest: 0.89055990 0.90379600 0.90579194 0.90936625 0.92817160
## ---------------------------------------------------------------------------
## b5
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18609      406    16848        1   0.2514  0.07687   0.1468   0.1652
##      .25      .50      .75      .90      .95
##   0.2011   0.2482   0.3031   0.3396   0.3598
##
## lowest : 0.08006774 0.08352830 0.08360645 0.08491613 0.08497245
## highest: 0.55240756 0.55939680 0.56861960 0.57122743 0.57889880
## ---------------------------------------------------------------------------
## b6
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18609      406    16863        1   0.1827  0.08793  0.07019  0.08663
##      .25      .50      .75      .90      .95
##  0.12622  0.17331  0.22305  0.29727  0.33825
##
## lowest : 0.03550000 0.03616667 0.03626785 0.03670000 0.03746371
## highest: 0.58307330 0.58979195 0.59474194 0.60415970 0.61979750
## ---------------------------------------------------------------------------
## b7
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18609      406    16810        1   0.1064  0.07094  0.03249  0.04020
##      .25      .50      .75      .90      .95
##  0.05883  0.08849  0.13167  0.20513  0.25483
##
## lowest : 0.01204138 0.01282143 0.01320000 0.01405893 0.01506290
## highest: 0.50048065 0.50056000 0.51021450 0.52273680 0.55494900
## ---------------------------------------------------------------------------
## EVI
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    17395     1620    15854        1   0.3024   0.1415   0.1212   0.1479
##      .25      .50      .75      .90      .95
##   0.2117   0.2825   0.3834   0.4834   0.5426
##
## lowest : 0.007224556 0.008124289 0.009238364 0.014921306 0.014974085
## highest: 0.730268500 0.735327700 0.738576300 0.739212300 0.758484800
## ---------------------------------------------------------------------------
## GCI
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18546      469    16926        1    2.908    1.752   0.6390   0.9973
##      .25      .50      .75      .90      .95
##   1.7987   2.7476   3.8548   4.9870   5.7303
##
## lowest :  0.004437923  0.006397247  0.007219228  0.010492164  0.012310988
## highest:  9.911120000  9.975493000 10.064940000 10.871304500 11.263188000
## ---------------------------------------------------------------------------
## NDVI
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18563      452    16928        1   0.5351   0.2271   0.1612   0.2197
##      .25      .50      .75      .90      .95
##   0.4054   0.5700   0.6912   0.7745   0.8116
##
## lowest : 0.002544436 0.005292194 0.005326130 0.005357474 0.007510734
```

```
## highest: 0.914746340 0.914772150 0.915391200 0.915637900 0.915831740
## ---------------------------------------------------------------------------
## NDWI
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18609      406    16984        1   0.1658   0.2224 -0.14604 -0.10814
##      .25      .50      .75      .90      .95
##  0.02816  0.16970  0.28347  0.37611  0.57355
##
## lowest : -0.2761985 -0.2730416 -0.2723996 -0.2689405 -0.2687093
## highest:  0.8205857  0.8271840  0.8343591  0.8380104  0.8547556
## ---------------------------------------------------------------------------
## NIRv
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18563      452    16928        1   0.1316  0.08005  0.03988  0.05013
##      .25      .50      .75      .90      .95
##  0.07697  0.11663  0.17518  0.23585  0.27508
##
## lowest : 0.001237986 0.001943630 0.002343579 0.002388369 0.002575980
## highest: 0.411172300 0.418155250 0.419231680 0.424213470 0.426510000
## ---------------------------------------------------------------------------
## kNDVI
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18609      406    16979        1   0.3057   0.2001  0.02509  0.04672
##      .25      .50      .75      .90      .95
##  0.16543  0.31532  0.44439  0.53669  0.57734
##
## lowest : 0.0000157 0.0000592 0.0000652 0.0000662 0.0000839
## highest: 0.6840645 0.6840799 0.6846792 0.6849552 0.6851109
## ---------------------------------------------------------------------------
## Percent_Snow
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18872      143     3426    0.618    11.62    19.92    0.000    0.000
##      .25      .50      .75      .90      .95
##    0.000    0.000    1.171   54.613   90.077
##
## lowest : 0.000000e+00 5.093379e-03 5.376344e-03 5.555556e-03 5.639098e-03
## highest: 9.988625e+01 9.989286e+01 9.990323e+01 9.993731e+01 1.000000e+02
## ---------------------------------------------------------------------------
## Fpar
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18080      935       86        1   0.4845   0.2297     0.14     0.19
##      .25      .50      .75      .90      .95
##     0.33     0.50     0.65     0.74     0.79
##
## lowest : 0.06 0.07 0.08 0.09 0.10, highest: 0.87 0.88 0.89 0.90 0.91
## ---------------------------------------------------------------------------
## Lai
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18080      935       63    0.998    1.431    1.214      0.2      0.3
##      .25      .50      .75      .90      .95
##      0.5      1.1      2.1      3.1      3.7
##
## lowest : 0.1 0.2 0.3 0.4 0.5, highest: 5.9 6.0 6.1 6.2 6.3
## ---------------------------------------------------------------------------
```

```
## LST_Day
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0     3279        1    289.8    15.47    266.7    271.7
##      .25      .50      .75      .90      .95
##    280.4    290.9    298.5    307.0    312.9
##
## lowest : 238.54 241.24 242.12 242.84 243.24, highest: 327.64 327.90 327.92 328.08 328.28
## ------------------------------------------------------------------------------
## LST_Night
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0     2375        1    277.2    10.87    260.0    264.5
##      .25      .50      .75      .90      .95
##    271.2    277.9    284.2    288.9    292.1
##
## lowest : 237.90 238.14 238.66 238.68 238.76, highest: 299.12 299.26 299.28 299.32 299.34
## ------------------------------------------------------------------------------
## MODIS_IGBP
##        n  missing distinct
##    19015        0       14
##
## lowest : BSV CRO CSH DBF EBF, highest: SAV URB WAT WET WSA
##
## Value        BSV   CRO   CSH   DBF   EBF   ENF   GRA    MF   OSH   SAV   URB
## Frequency      8  4197    99  1576   306  1973  2145  2032   894  2475    73
## Proportion 0.000 0.221 0.005 0.083 0.016 0.104 0.113 0.107 0.047 0.130 0.004
##
## Value        WAT   WET   WSA
## Frequency    452   173  2612
## Proportion 0.024 0.009 0.137
## ------------------------------------------------------------------------------
## MODIS_PFT
##        n  missing distinct
##    19015        0        9
##
## lowest : CRO   DBF   EBF   ENF   GRA  , highest: GRA   MF    Other SA    SH
##
## Value        CRO   DBF   EBF   ENF   GRA    MF Other    SA    SH
## Frequency   4197  1576   306  1973  2145  2032   706  5087   993
## Proportion 0.221 0.083 0.016 0.104 0.113 0.107 0.037 0.268 0.052
## ------------------------------------------------------------------------------
## koppen_sub
##        n  missing distinct
##    19015        0       21
##
## lowest : Af  Am  Aw  BSh BSk, highest: Dsc Dwa Dwb Dwc ET
## ------------------------------------------------------------------------------
## koppen
##        n  missing distinct
##    19015        0        5
##
## lowest : Arid      Cold      Polar     Temperate Tropical
## highest: Arid      Cold      Polar     Temperate Tropical
##
## Value           Arid      Cold     Polar Temperate  Tropical
```

```
## Frequency       1957     11013       299      5102       644
## Proportion     0.103     0.579     0.016     0.268     0.034
## ---------------------------------------------------------------------------
## CO2_concentration
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    19015        0      238        1    390.3    12.92    373.0    375.7
##      .25      .50      .75      .90      .95
##    381.5    389.3    398.2    406.3    409.9
##
## lowest : 369.070 369.330 369.415 369.560 369.745
## highest: 412.435 412.715 412.780 413.135 413.325
## ---------------------------------------------------------------------------
```