Applied Geodata Science I

Session 3

Prof. Dr. Benjamin Stocker

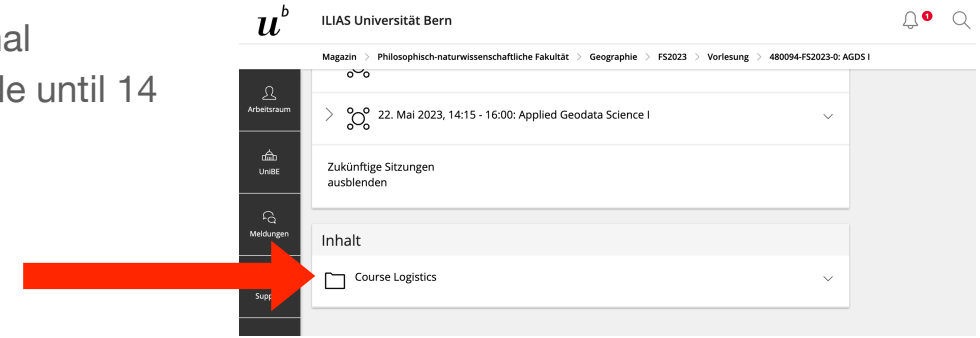
Spring semester 2023

GECO

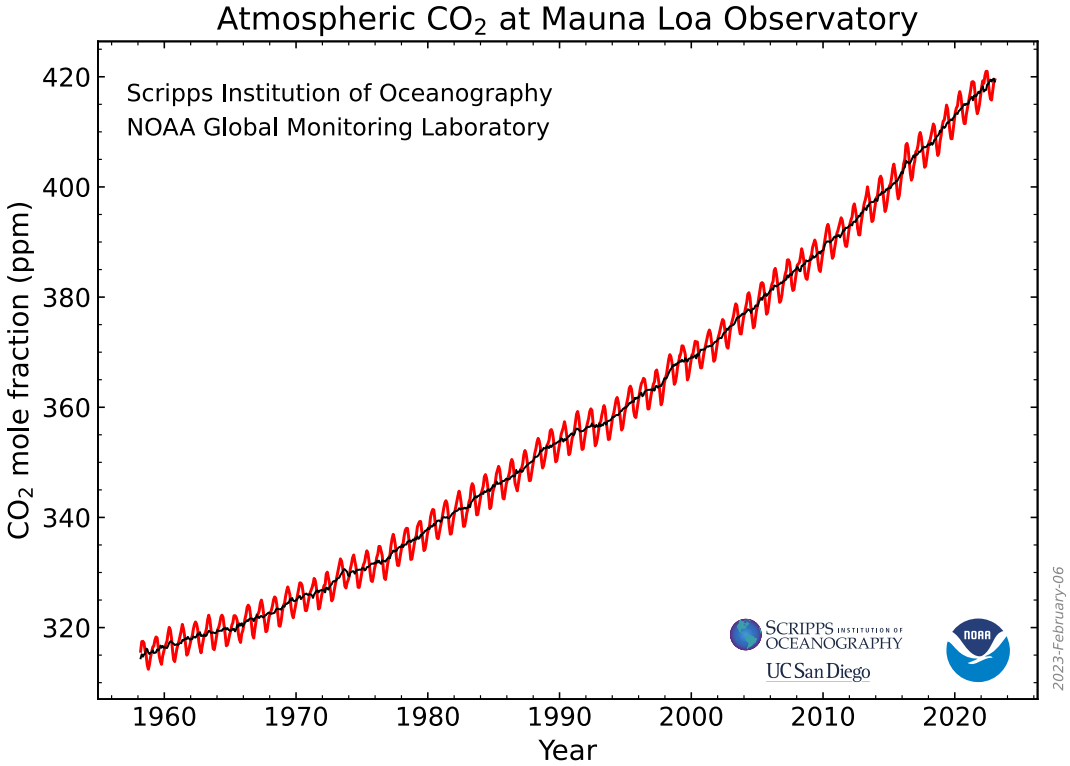
UNIVERSITÄT BERN

Updated info on course logistics

- On ILIAS
- AGDS I is required for the Proseminar in Applied Geodata Science, held in the autumn semester after AGDS I.
- The course is not formally “creditable” (German: anrechenbar) towards the Master studies in Geography. Requests for crediting can be made and will be handled on an individual basis.
- Master students in Climate Sciences can credit AGDS I as Elective Module.
- Definite course registrations have to be made until 15.03.2023.
- Registrations for the exam (in the form of the Final Report, see separate document) have to be made until 14 days before the last session (22.05.2023).

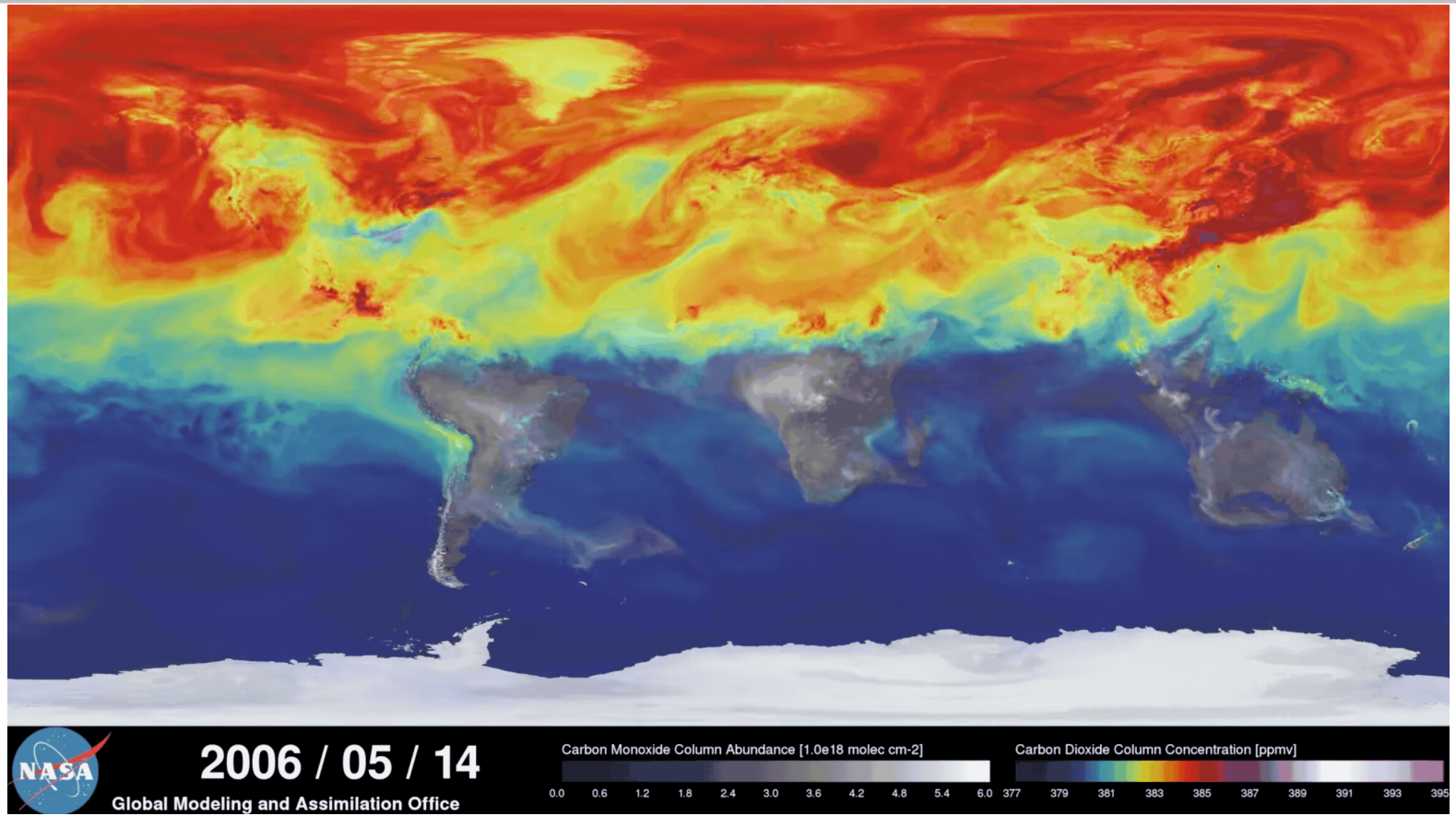


Atmospheric CO2 concentration



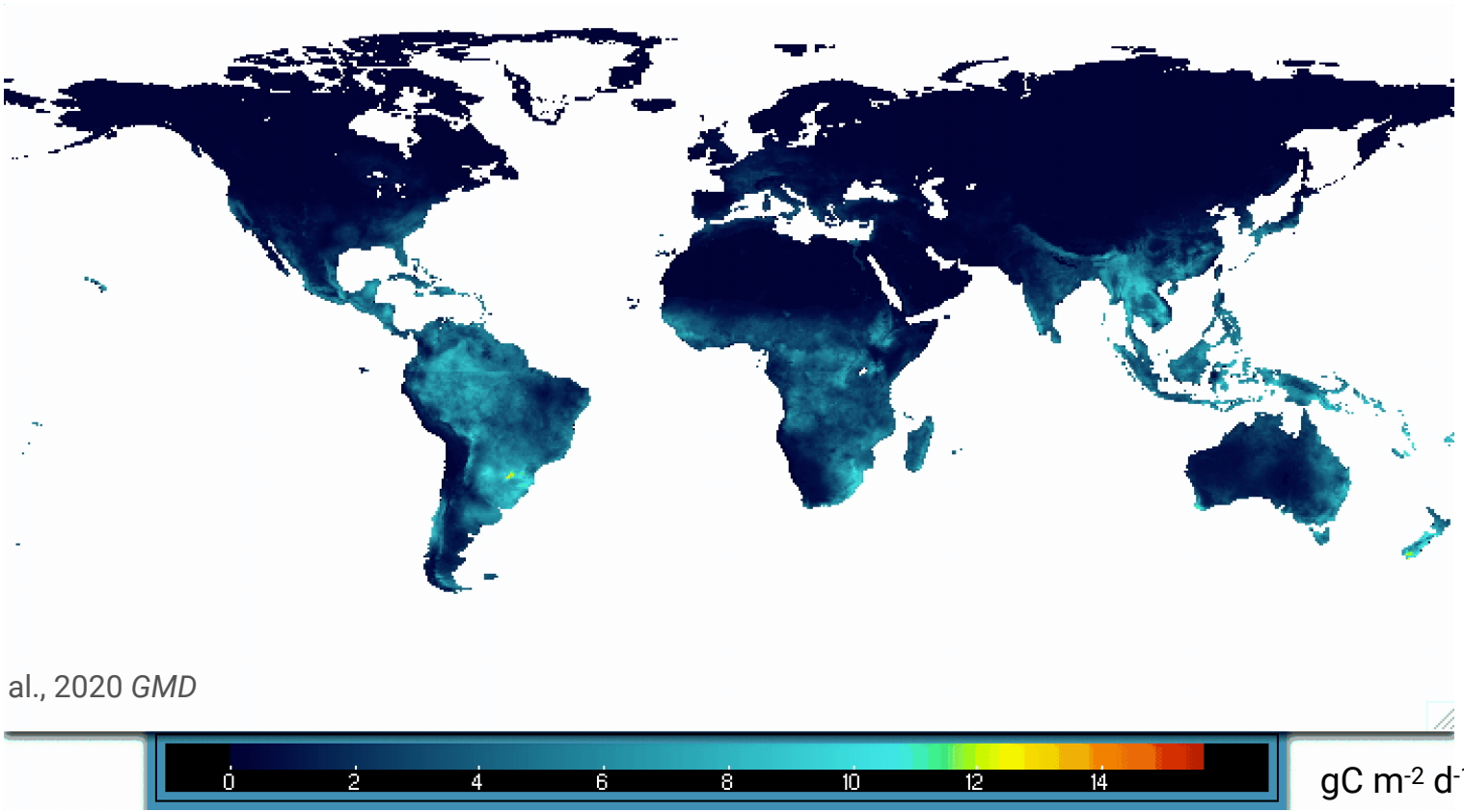
https://gml.noaa.gov/ccgg/trends/

Atmospheric CO2 concentration



Source: NASA (https://www.youtube.com/watch?v=x1SgmFa0r04)

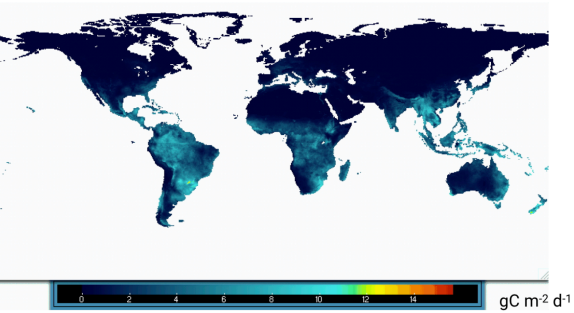
Seasonality in gross primary production



Stocker et al., 2020 *GMD*

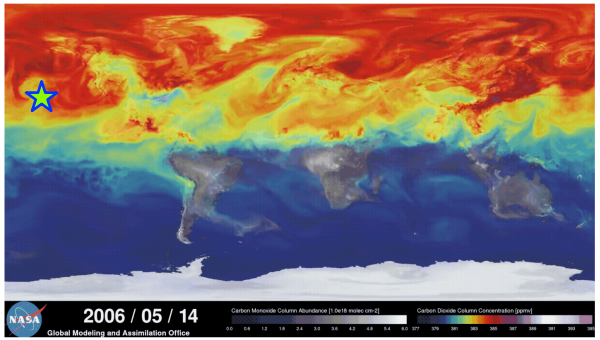
Breathing of the Earth

Photosynthetic CO_2 assimilation

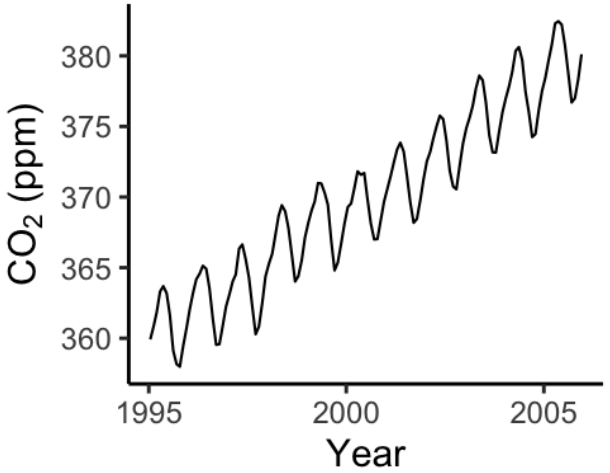


Stocker et al., 2020 *GMD*

Atmospheric CO_2 concentration

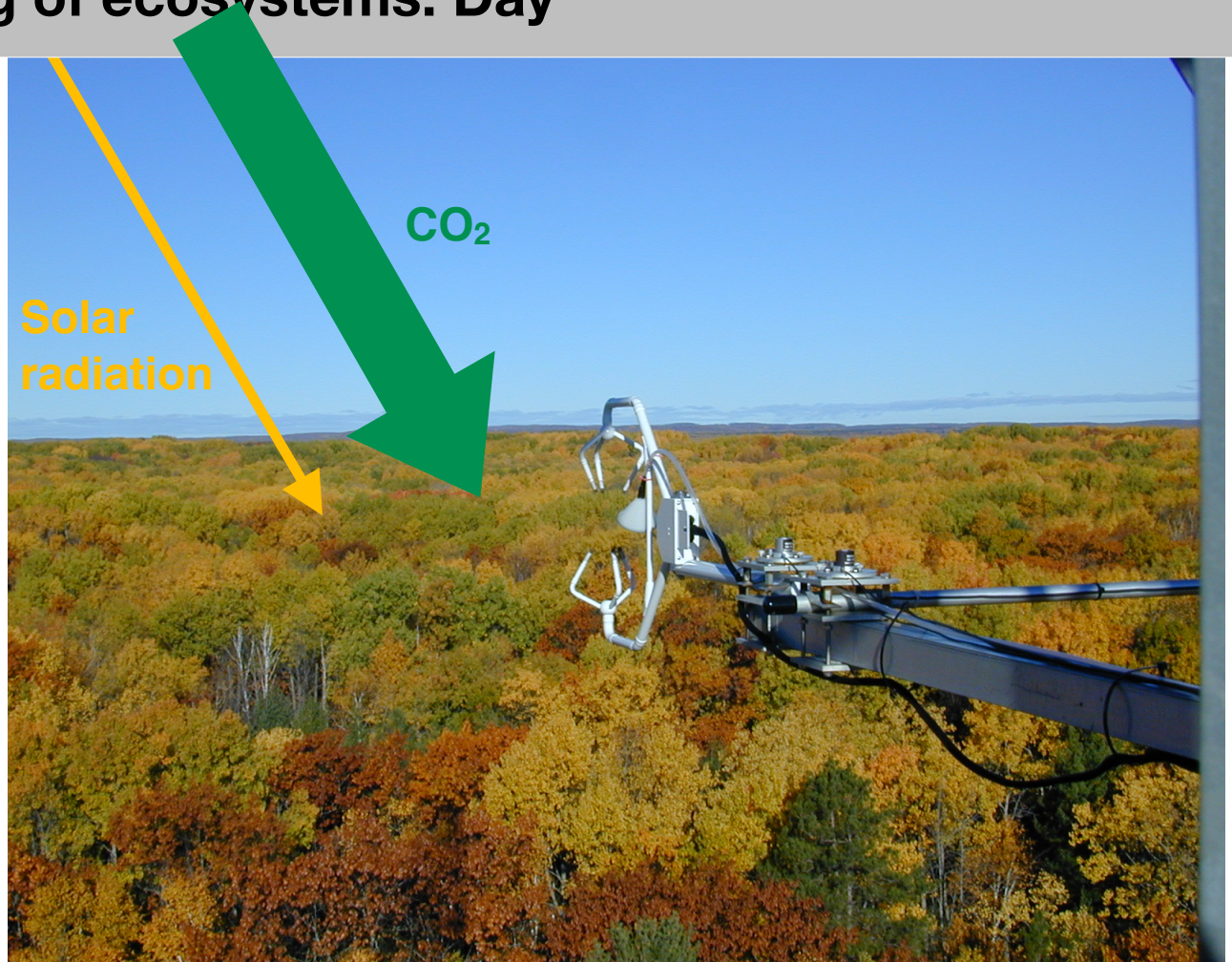


CO_2 concentration at Mauna Loa



Keeling et al., 1976 *Tellus*
Thoning et al., 1989 *JGR*

Breathing of ecosystems: Day

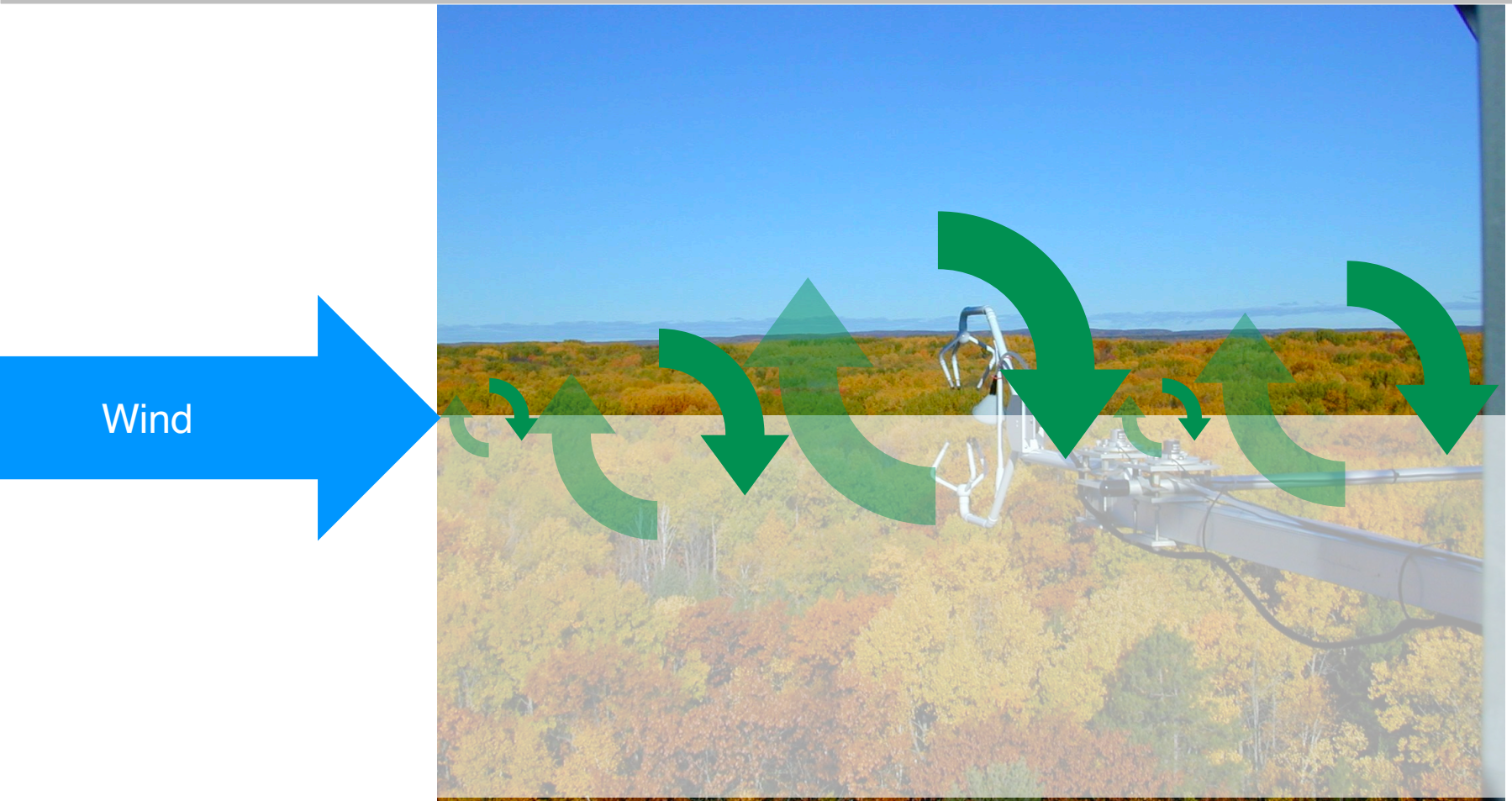


Breathing of ecosystems: Night

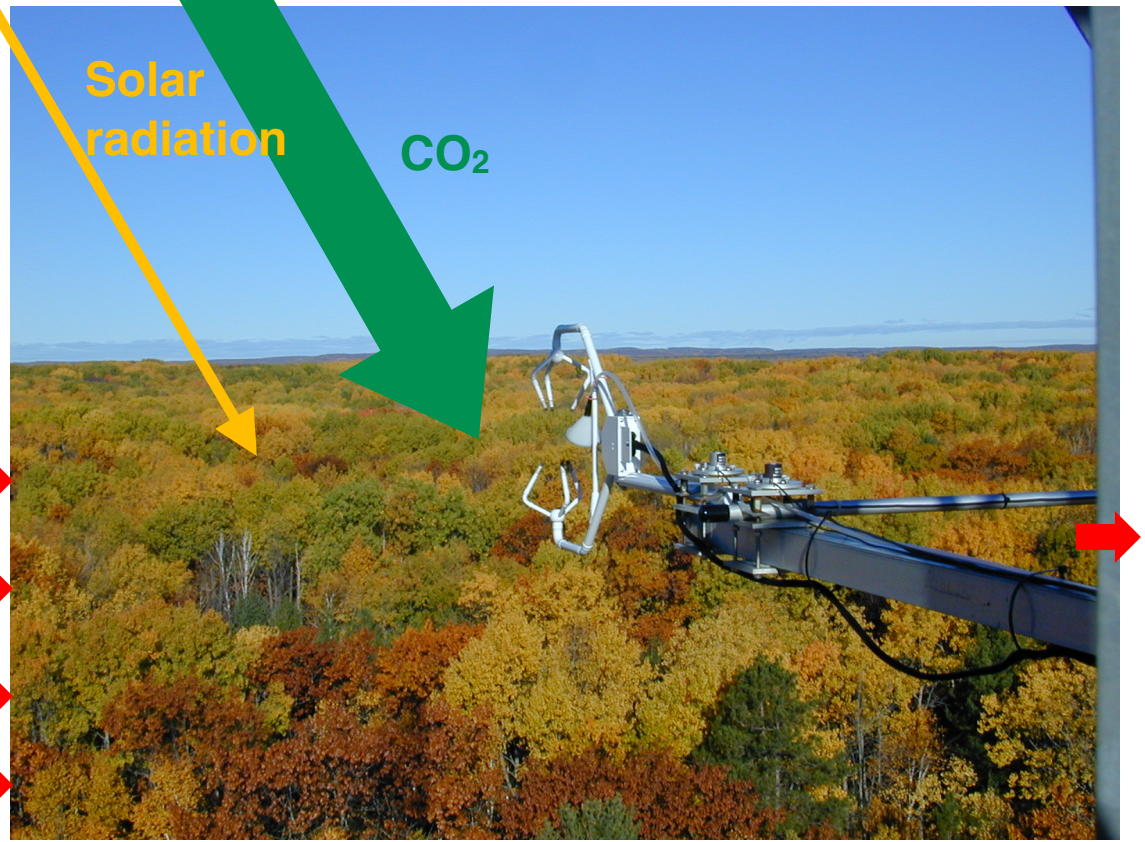
https://andrewsforest.oregonstate.edu/research/programs

Land-atmosphere gas exchange through eddies

Eddies and CO_2 transport



Breathing of ecosystems



Temperature

Atmospheric humidity

Soil moisture

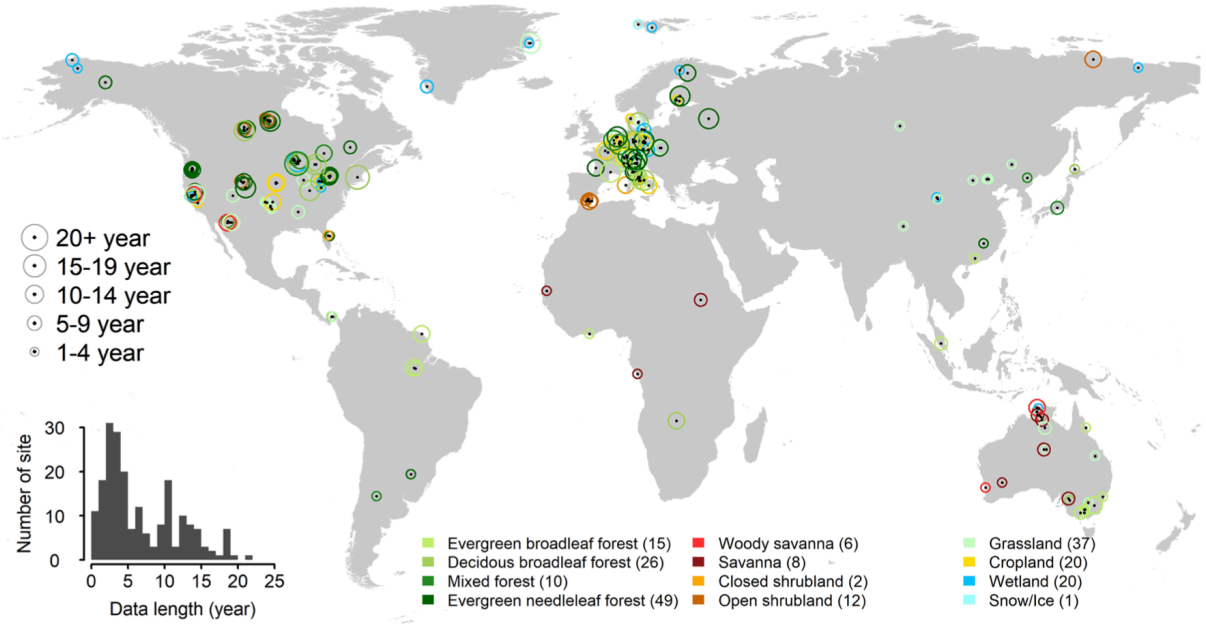
Leaf area

Ecosystem respiration

Net CO_2 uptake

Gross primary production

The FLUXNET network



Pastorello et al., 2020 *Sci. Dat.*

SwissFluxNet

Image Markus Staudinger

Lägern
CH-Lae

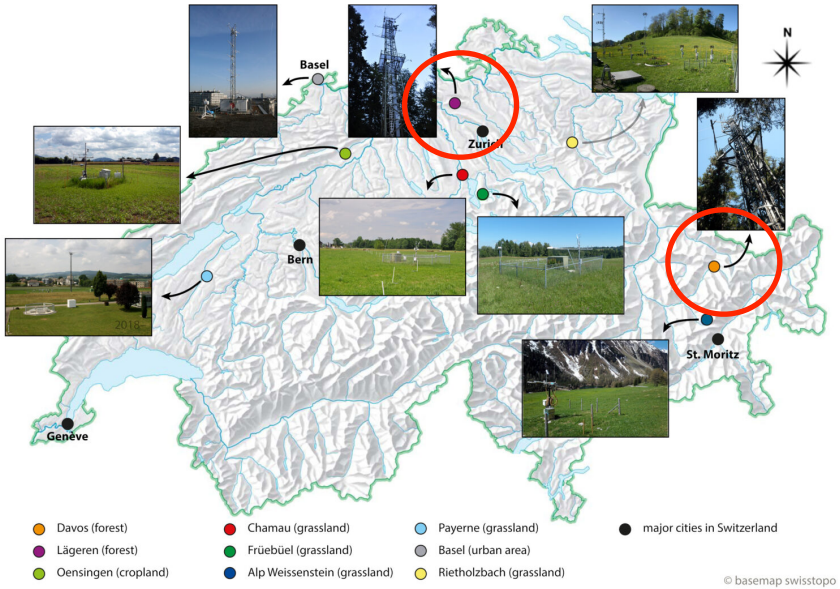


Image Lukas Hörtnagl

Davos
CH-Dav

https://www.swissfluxnet.ethz.ch/

Variables in a FLUXNET-standard dataset

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

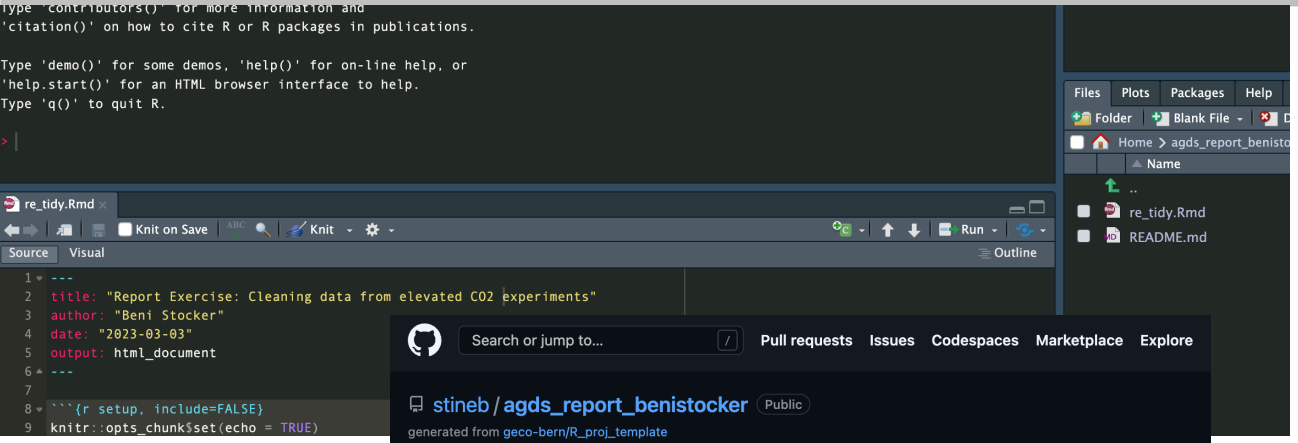
DATA DESCRIPTOR

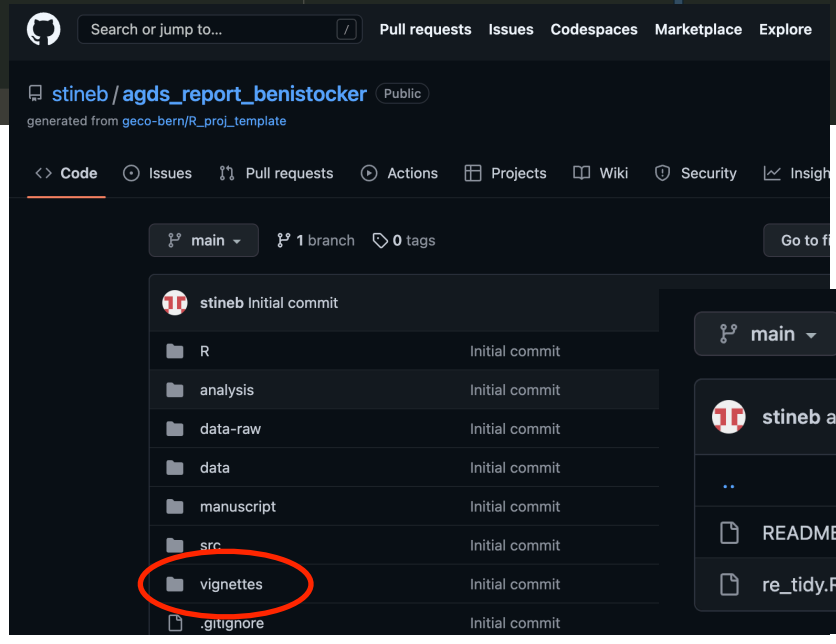
The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data

Gilberto Pastorello *et al.*#

- `TIMESTAMP_START` : Hour and day of the start of the measurement period for which the respective row's data are representative. Provided in a format of "YYYYMMDDhhmm".
- `TIMESTAMP_END` : Hour and day of the end of the measurement period for which the respective row's data are representative. Provided in a format of "YYYYMMDDhhmm".
- `TA_*` (°C): Air temperature.
- `SW_IN_*` (W m^{-2}): Shortwave incoming radiation
- `LW_IN_*` (W m^{-2}): Longwave incoming radiation
- `VPD_*` (hPa): Vapor pressure deficit (the difference between actual and saturation water vapor pressure)
- `PA_*` (kPa): Atmospheric pressure
- `P_*` (mm): Precipitation
- `WS_*` (m $^{-1}$): Wind speed
- `SWC_*` (%): Volumetric soil water content
- `GPP_*` (μmol CO_2 m^{-2} s^{-1}): Gross primary production (the ecosystem-level gross CO_2 uptake flux driven by photosynthesis)
- `*_QC` : Quality control information for the variable `*` . Important for us: `NEE_*_QC` is the quality control information for the net ecosystem CO_2 exchange flux (`NEE_*`) and for GPP derived from the corresponding NEE estimate (`GPP_*`). 0 = measured, 1 = good quality gap-filled, 2 = medium, 3 = poor.

Report Exercise

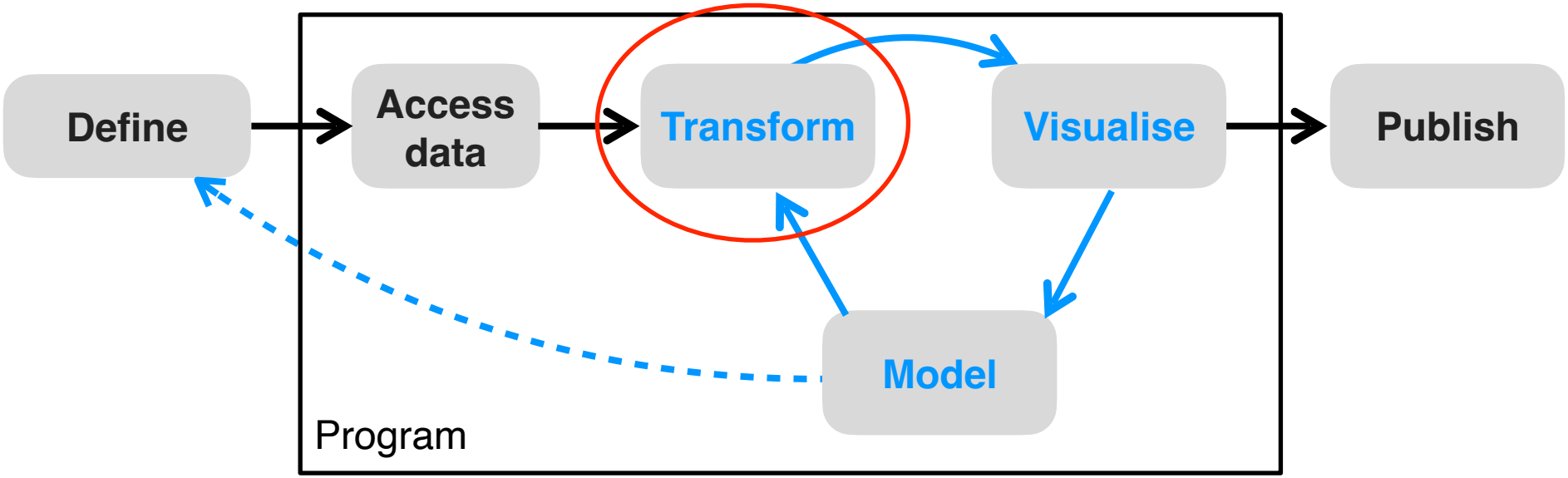




Report Exercise: Criteria

File on ILIAS

The data science workflow



Overview

Chapters 1 and 2

- Getting ready with R
- Data frames

ncol() →

nrow() ↓

```
> hhdf
# A tibble: 192,864 x 20
   TIMESTAMP_START     TIMESTAMP_END       TA_F SW_IN_F LW_IN_F VPD_F  PA_F   P_F  WS_F CO2_F_MDS PPFD_IN
   <dttm>              <dttm>             <dbl>   <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl>      <dbl>   <dbl>
 1 2004-01-01 00:00:00 2004-01-01 00:30:00    NA      NA    304.    NA  93.3 0.014    NA         NA      NA
 2 2004-01-01 00:30:00 2004-01-01 01:00:00    NA      NA    304.    NA  93.3 0.014    NA         NA      NA
 3 2004-01-01 01:00:00 2004-01-01 01:30:00    NA      NA    281.    NA  93.3 0        NA         NA      NA
 4 2004-01-01 01:30:00 2004-01-01 02:00:00    NA      NA    281.    NA  93.3 0        NA         NA      NA
 5 2004-01-01 02:00:00 2004-01-01 02:30:00    NA      NA    281.    NA  93.3 0        NA         NA      NA
 6 2004-01-01 02:30:00 2004-01-01 03:00:00    NA      NA    281.    NA  93.3 0        NA         NA      NA
 7 2004-01-01 03:00:00 2004-01-01 03:30:00    NA      NA    281.    NA  93.2 0        NA         NA      NA
 8 2004-01-01 03:30:00 2004-01-01 04:00:00    NA      NA    281.    NA  93.2 0        NA         NA      NA
 9 2004-01-01 04:00:00 2004-01-01 04:30:00    NA      NA    264.    NA  93.2 0        NA         NA      NA
10 2004-01-01 04:30:00 2004-01-01 05:00:00    NA      NA    264.    NA  93.2 0        NA         NA      NA
# … with 192,854 more rows, and 9 more variables: GPP_NT_VUT_REF <dbl>, SWC_F_MDS_1 <dbl>, SWC_F_MDS_2 <dbl>,
#   SWC_F_MDS_3 <dbl>, WS <dbl>, WD <dbl>, RH <dbl>, NIGHT <dbl>, NEE_VUT_REF_QC <dbl>
```

Chapter 3

- Understand data and its “dimensions” and “variations”
- Data transformation
- Entering the R *tidyverse*

```
co2_concentration_monthly |>
  pivot_longer(cols = 2:13, names_to = "month", values_to = "co2")
```

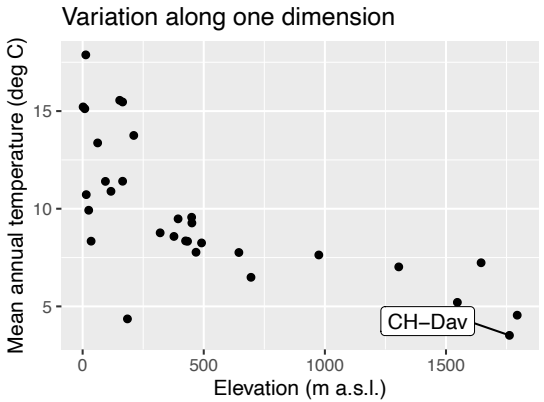
```
## # A tibble: 36 × 3
##     year month   co2
##    <int> <chr> <dbl>
##  1  1959 Jan    315.
##  2  1959 Feb    316.
##  3  1959 Mar    316.
```

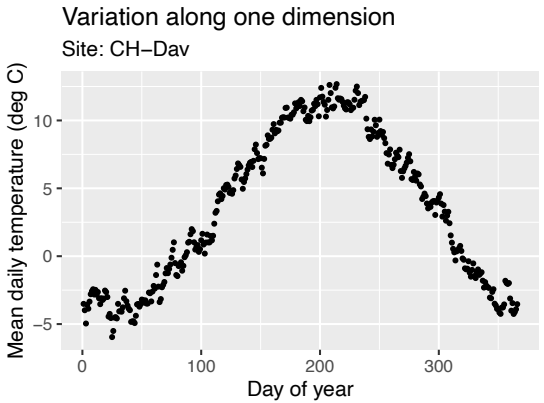
Variations and dimensions

- **Variation**: How data points vary between observations.
- **Dimension**: Factors that vary along with observations.

Example:

- **Variation along 1 dimension**:
  - Mean annual temperature along an elevational gradient.
  - Temperature mean seasonality (of one site)



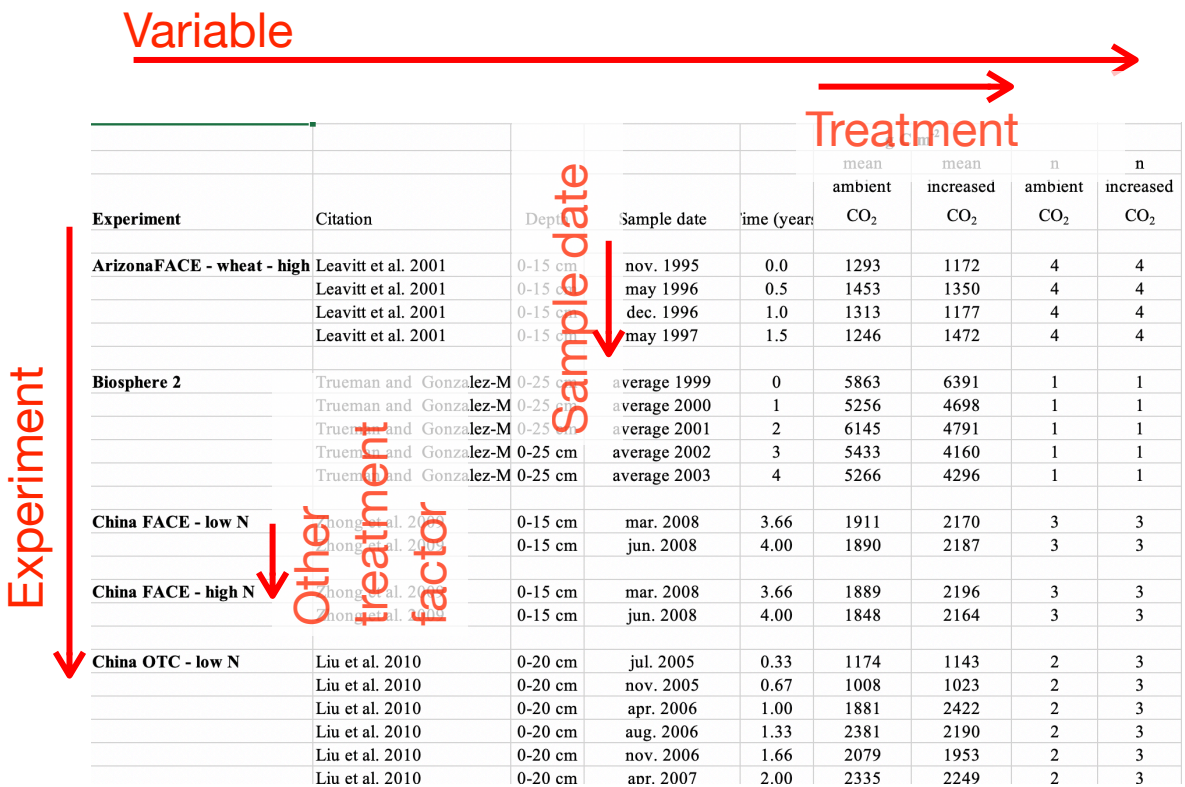


Variations and dimensions

- **Variation along >1 dimensions**:
  - Mean seasonality along an elevation gradient.
  - Data for experiments meta-analysis

➔ **Data organisation?**

➔ **Data visualisation? (Chapter 4)**



Experiment	Citation	Depth	Sample date	Time (years)	mean ambient CO_2	mean increased CO_2	n ambient CO_2	n increased CO_2
ArizonaFACE - wheat - high	Leavitt et al. 2001	0-15 cm	nov. 1995	0.0	1293	1172	4	4
	Leavitt et al. 2001	0-15 cm	may 1996	0.5	1453	1350	4	4
	Leavitt et al. 2001	0-15 cm	dec. 1996	1.0	1313	1177	4	4
	Leavitt et al. 2001	0-15 cm	may 1997	1.5	1246	1472	4	4
Biosphere 2	Trueman and Gonzalez-M	0-25 cm	average 1999	0	5863	6391	1	1
	Trueman and Gonzalez-M	0-25 cm	average 2000	1	5256	4698	1	1
	Trueman and Gonzalez-M	0-25 cm	average 2001	2	6145	4791	1	1
	Trueman and Gonzalez-M	0-25 cm	average 2002	3	5433	4160	1	1
	Trueman and Gonzalez-M	0-25 cm	average 2003	4	5266	4296	1	1
China FACE - low N	Zhong et al. 2017	0-15 cm	mar. 2008	3.66	1911	2170	3	3
	Zhong et al. 2017	0-15 cm	jun. 2008	4.00	1890	2187	3	3
China FACE - high N	Zhong et al. 2017	0-15 cm	mar. 2008	3.66	1889	2196	3	3
	Zhong et al. 2017	0-15 cm	jun. 2008	4.00	1848	2164	3	3
China OTC - low N	Liu et al. 2010	0-20 cm	jul. 2005	0.33	1174	1143	2	3
	Liu et al. 2010	0-20 cm	nov. 2005	0.67	1008	1023	2	3
	Liu et al. 2010	0-20 cm	apr. 2006	1.00	1881	2422	2	3
	Liu et al. 2010	0-20 cm	aug. 2006	1.33	2381	2190	2	3
	Liu et al. 2010	0-20 cm	nov. 2006	1.66	2079	1953	2	3
	Liu et al. 2010	0-20 cm	apr. 2007	2.00	2335	2249	2	3

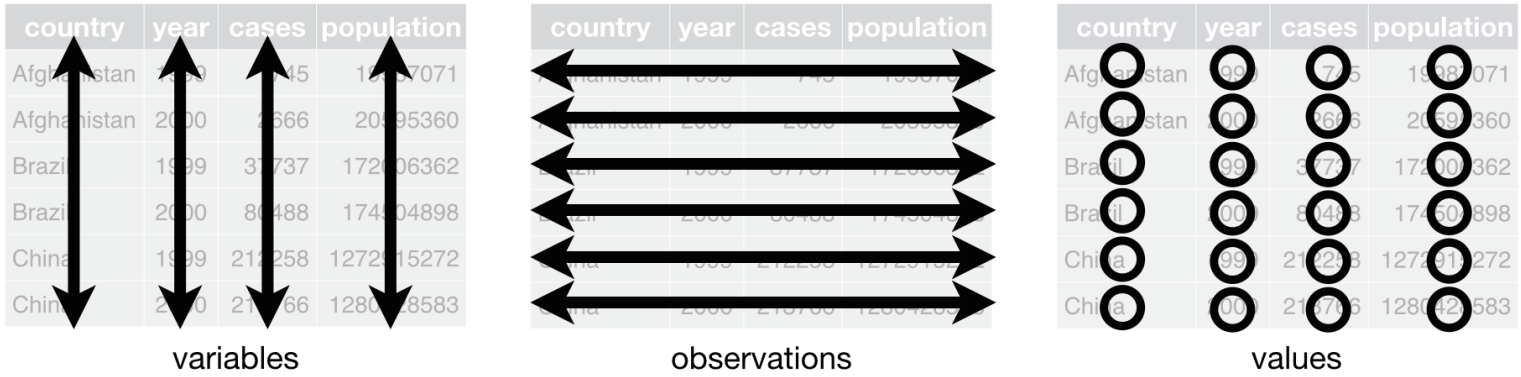
Tidy data

Tidy?

```
## # A tibble: 3 × 13
##    year   Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
##   <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  1959  315.  316.  316.  318.  318.  318   316.  315.  314.  313.  315.  315.
## 2  1960  316.  317.  317.  319.  320.  319.  318.  316.  314   314.  315.  316.
## 3  1961  317.  318.  318.  319.  320.  320.  318.  317.  315.  315.  316.  317.
```

```
## # A tibble: 36 × 3
##     year month   co2
##    <int> <chr> <dbl>
##  1  1959 Jan    315.
##  2  1959 Feb    316.
##  3  1959 Mar    316.
##  4  1959 Apr    318.
##  5  1959 May    318.
##  6  1959 Jun    318
##  7  1959 Jul    316.
##  8  1959 Aug    315.
##  9  1959 Sep    314.
## 10  1959 Oct    313.
## # … with 26 more rows
```

Tidy data



Wickham & Grolemund, 2017