Applied Geodata Science 1

Session 3

Prof. Dr. Benjamin Stocker & Dr. Fabian Bernhard

Spring semester 2025

GECO

UNIVERSITÄT BERN

AGDS I logistics

Find this information on:
https://geco-bern.github.io/agds1_course/

1 Getting started 17.02.

2 Programming primers 24.02.

3 Data wrangling 03.03.
- Report Exercise: Tidy data

4 Data visualisation 10.03.
- Report Exercise: Air quality data

5 Data variety 17.03.

6 Open Science practice 24.03.

7 Code management 31.03.
- Report Exercise: Collaborating with git

8 CARE SESSION I 07.04.
- Work on R. Ex. 7 **as team exercise!**

9 Regression 14.04.
- Report Exercise: Stepwise regression

10 Supervised ML I 28.04.
- Report Exercise: KNN

11 Supervised ML II 05.05.
- Report Exercise: Flux modelling

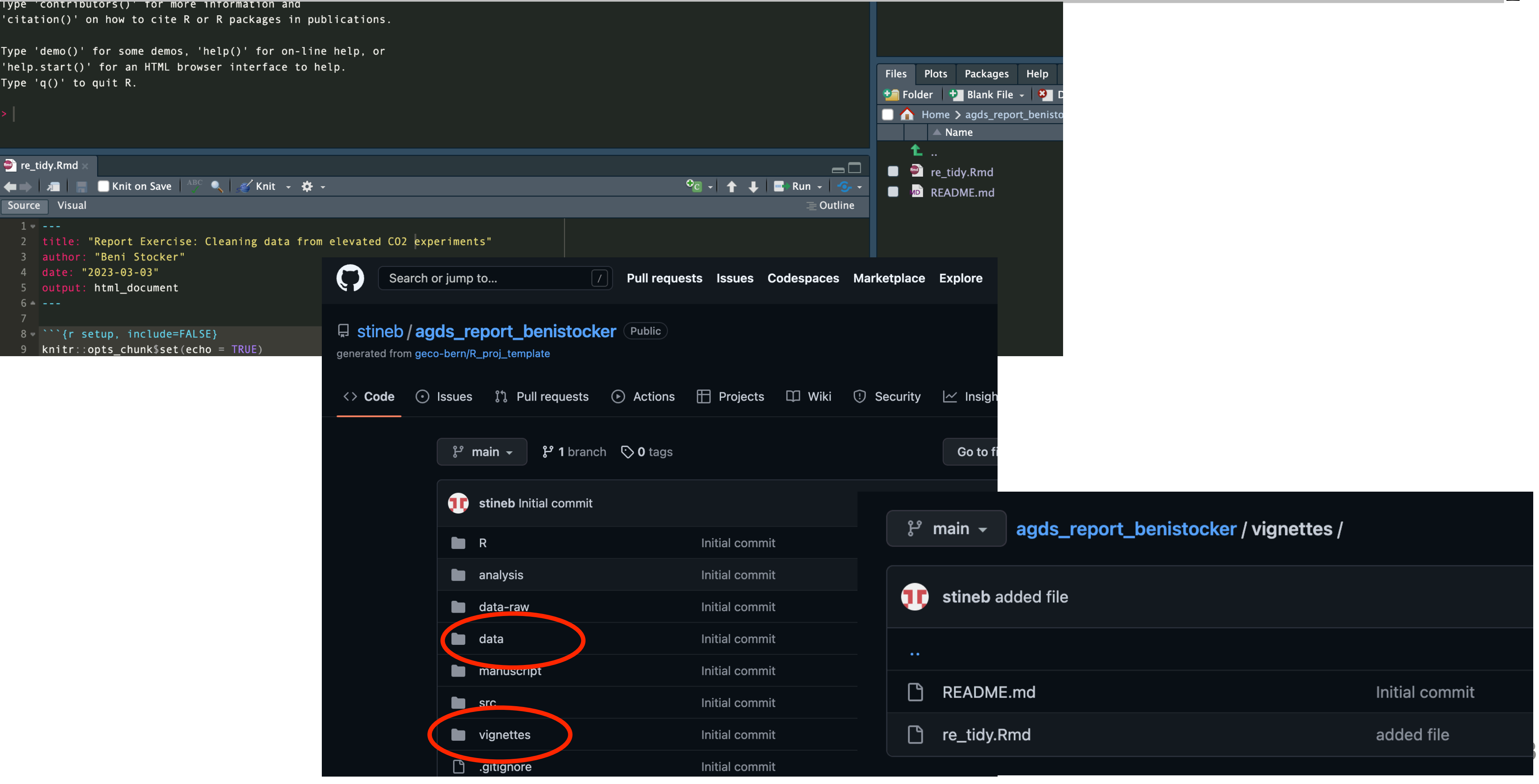
12 Random Forest 12.05.

13 Interpretable ML 19.05.

14 CARE SESSION II 26.05.
- Catch-up and preparations for final report

FINAL REPORT
- Hand in as reproducible code (*git* repository)
- Repo 1: Consists of 5 elements implementing Report Exercises (separate RMarkdown files)
- Repo 2 and 3: From Report Ex. 7

Report Exercise: Repo 1 "agds_report_yourname"

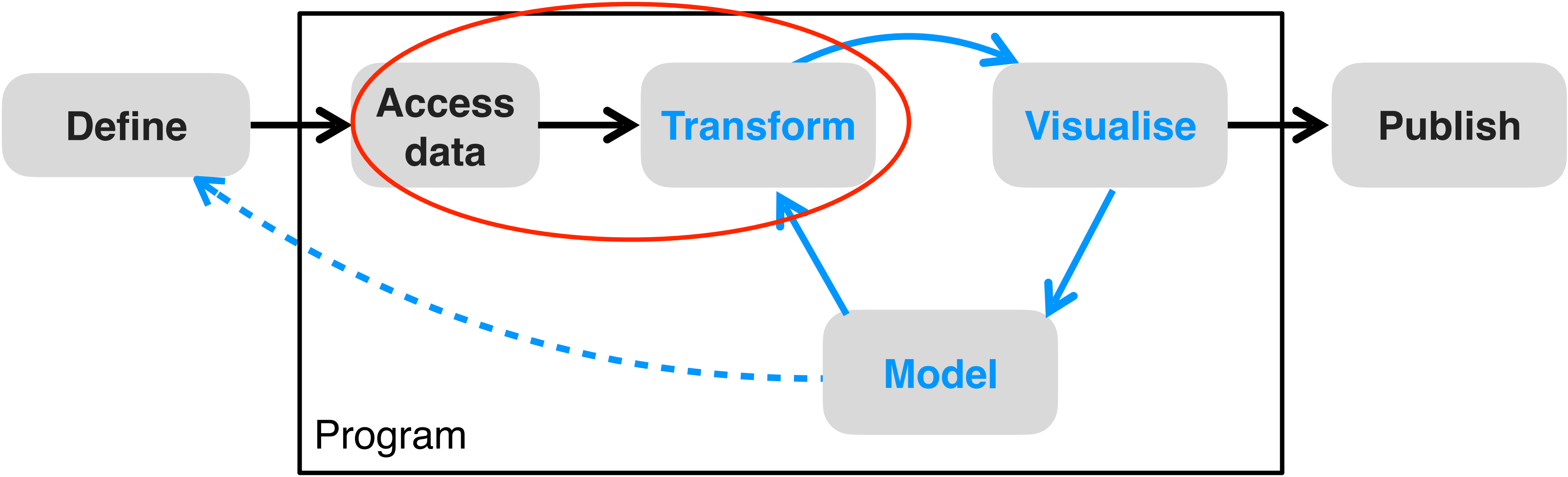


Report Exercise: Grading criteria

Criteria reported at:
https://geco-bern.github.io/agds1_course/logistics.html
or file on ILIAS:

- reproducible code
- good coding practice (legibility, workspace organisation, code mgmt)
- quality of visualisations
- interpretations

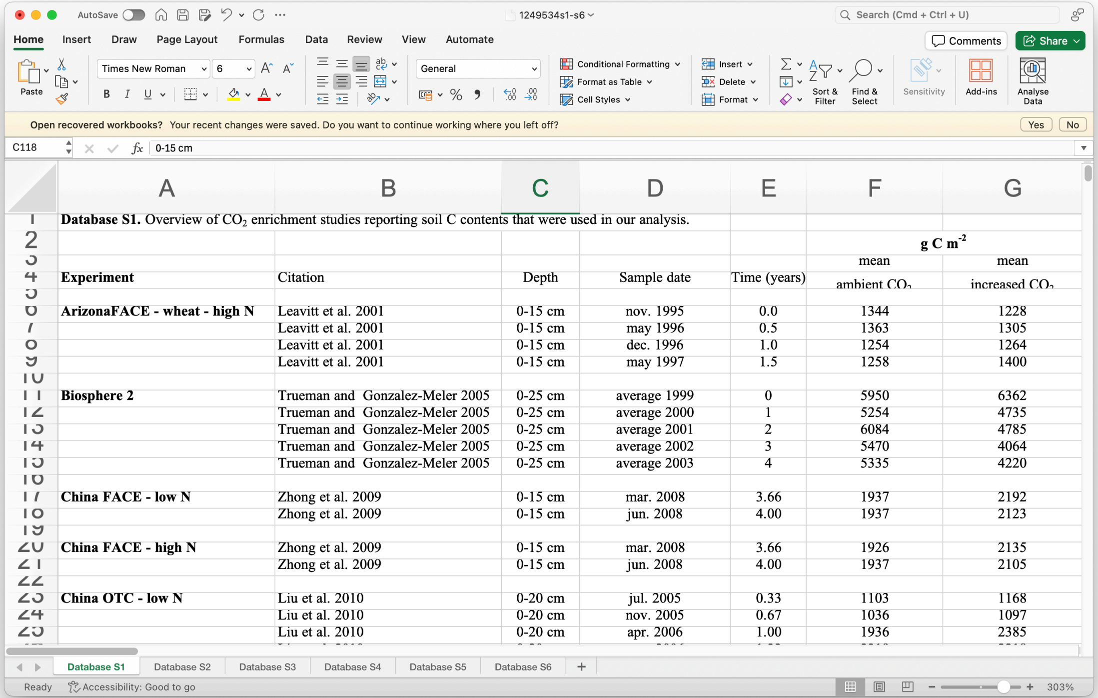
The data science workflow



A table: has rows and columns

An example of a shared Excel file:
- the file has 2 dimensions
- how many has the data?

How to:
- make a table in an article
- make a table for sharing data
- (make a graph)



Database S1. Overview of CO_2 enrichment studies reporting soil C contents that were used in our analysis.

Experiment	Citation	Depth	Sample date	Time (years)	mean ambient CO_2 ($g\ C\ m^{-2}$)	mean increased CO_2 ($g\ C\ m^{-2}$)
ArizonaFACE - wheat - high N	Leavitt et al. 2001	0-15 cm	nov. 1995	0.0	1344	1228
	Leavitt et al. 2001	0-15 cm	may 1996	0.5	1363	1305
	Leavitt et al. 2001	0-15 cm	dec. 1996	1.0	1254	1264
	Leavitt et al. 2001	0-15 cm	may 1997	1.5	1258	1400
Biosphere 2	Trueman and Gonzalez-Meler 2005	0-25 cm	average 1999	0	5950	6362
	Trueman and Gonzalez-Meler 2005	0-25 cm	average 2000	1	5254	4735
	Trueman and Gonzalez-Meler 2005	0-25 cm	average 2001	2	6084	4785
	Trueman and Gonzalez-Meler 2005	0-25 cm	average 2002	3	5470	4064
	Trueman and Gonzalez-Meler 2005	0-25 cm	average 2003	4	5335	4220
China FACE - low N	Zhong et al. 2009	0-15 cm	mar. 2008	3.66	1937	2192
	Zhong et al. 2009	0-15 cm	jun. 2008	4.00	1937	2123
China FACE - high N	Zhong et al. 2009	0-15 cm	mar. 2008	3.66	1926	2135
	Zhong et al. 2009	0-15 cm	jun. 2008	4.00	1937	2105
China OTC - low N	Liu et al. 2010	0-20 cm	jul. 2005	0.33	1103	1168
	Liu et al. 2010	0-20 cm	nov. 2005	0.67	1036	1097
	Liu et al. 2010	0-20 cm	apr. 2006	1.00	1936	2385

Depends what you want to show:
Which differences do you want to showcase?
To move around between formats => **Transformation => tidyverse packages**

Overview

`ncol()`

Chapters 1 and 2

- Getting ready with R
- Data frames

`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")
```

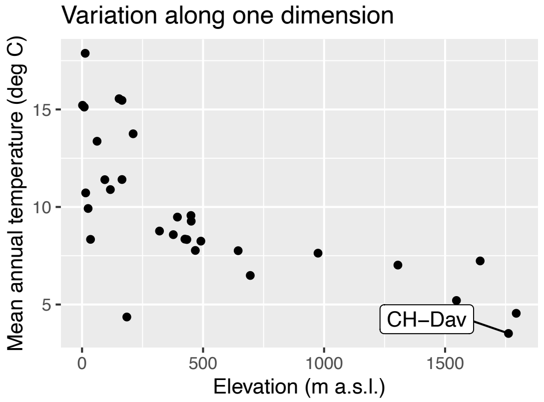
```
## # A tibble: 36 × 3
##      year month   co2
##     <int> <chr> <dbl>
##   1  1959 Jan    315.
##   2  1959 Feb    316.
##   3  1959 Mar    316.
```

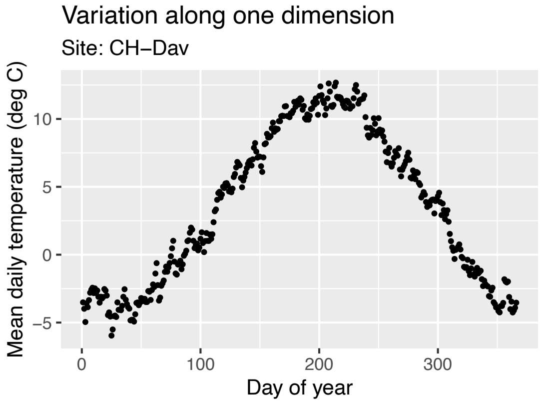
Variations and dimensions

- **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)
- **Variation along >1 dimensions**:
  - Temperature mean seasonality along an elevation gradient.



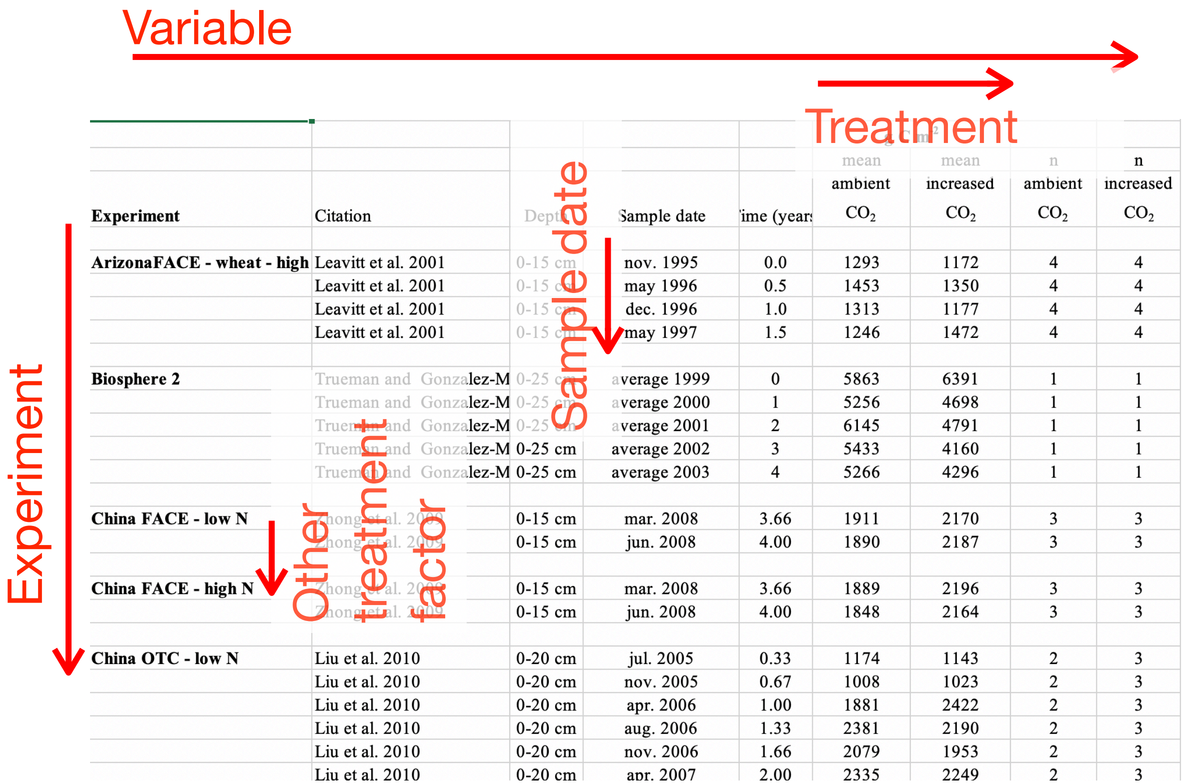


Variations and dimensions

- **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. 2009	0-15 cm	mar. 2008	3.66	1911	2170	3	3
	Zhong et al. 2009	0-15 cm	jun. 2008	4.00	1890	2187	3	3
China FACE - high N	Zhong et al. 2009	0-15 cm	mar. 2008	3.66	1889	2196	3	3
	Zhong et al. 2009	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

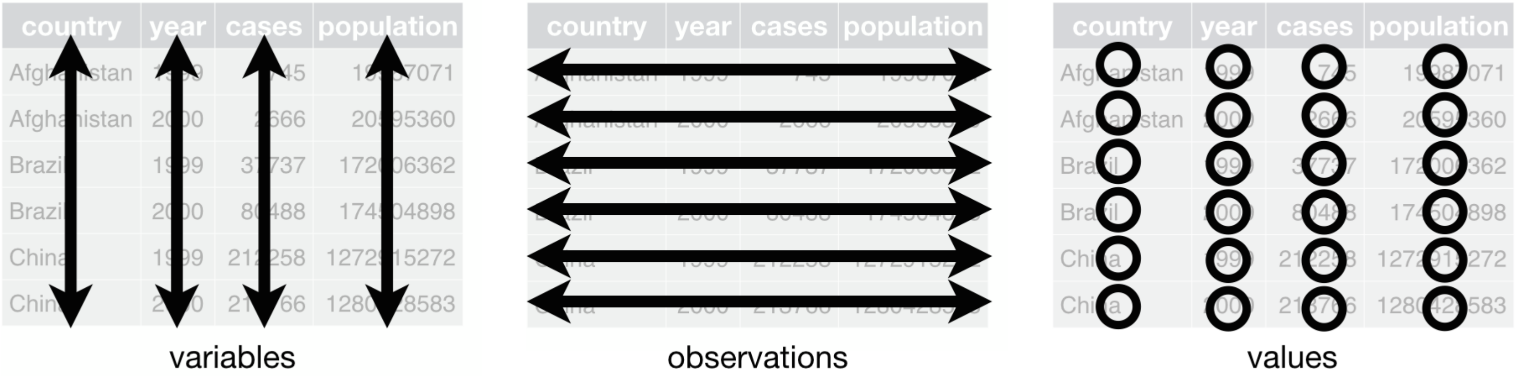
Tidy data

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

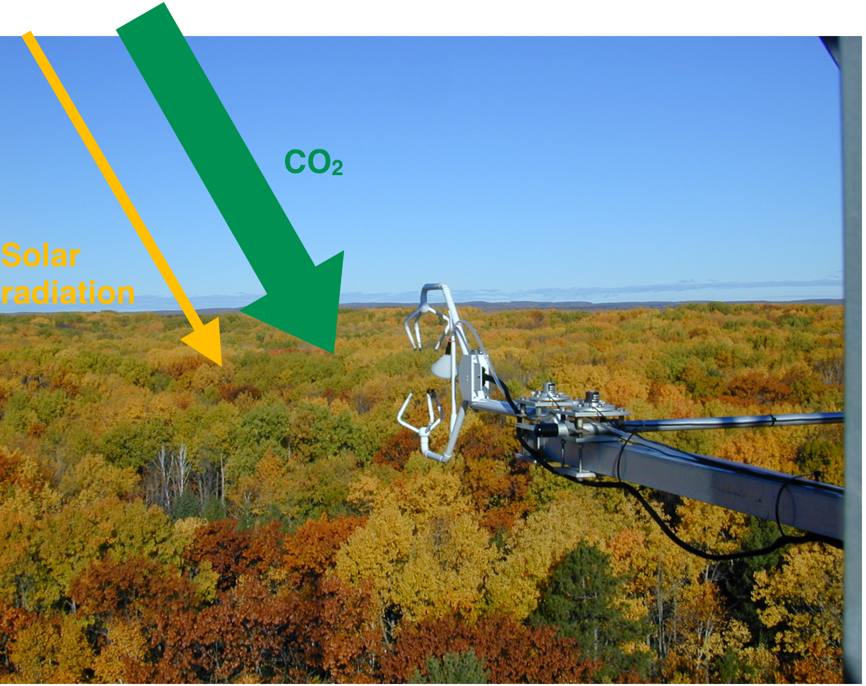


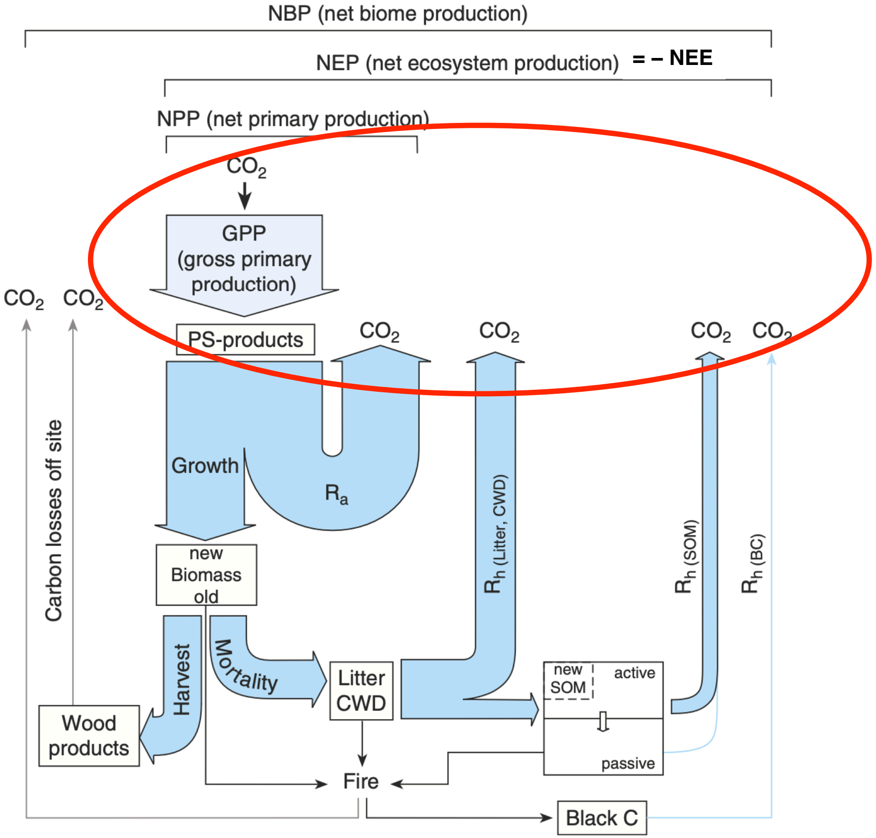
There are three interrelated rules which make a dataset tidy:

1. Each variable must have its own column.
2. Each observation must have its own row.
3. Each value must have its own cell.

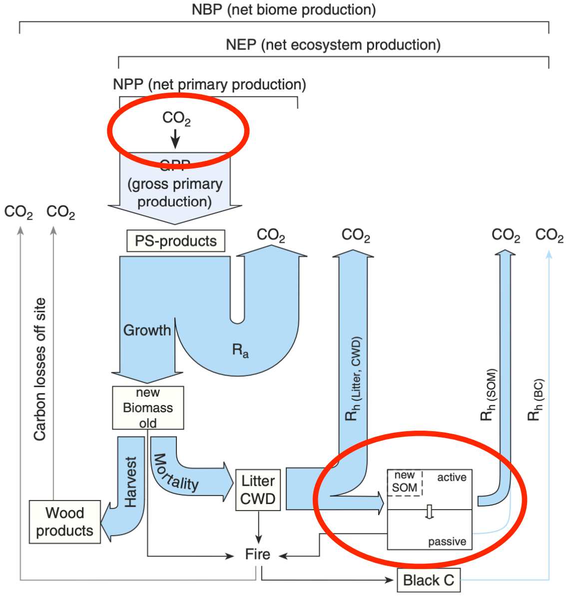
Wickham & Grolemund, 2017

Exercises Chapter 3: CO2 exchange land-atmosphere





Report Exercise Chapter 3



Science

Current Issue | First release papers | Archive | About

HOME > SCIENCE > VOL. 344, NO. 6183 > FASTER DECOMPOSITION UNDER INCREASED ATMOSPHERIC CO_2 LIMITS SOIL CARBON STORAGE

REPORT

Faster Decomposition Under Increased Atmospheric CO_2 Limits Soil Carbon Storage

KEES JAN VAN GROENIGEN, XUAN QI, CRAIG W. OSENBERG, YIQI LUO, AND BRUCE A. HUNGATE Authors Info & Affiliations

SCIENCE • 24 Apr 2014 • Vol 344, Issue 6183 • pp. 508-509 • DOI: 10.1126/science.1249534

Increasing Turnover

Rising concentrations of CO_2 in the atmosphere are likely to stimulate plant growth, leading to increased biomass and soil carbon stocks, thereby lessening the rate of the rise in atmospheric CO_2 levels. **Van Groenigen *et al.*** (p. 508, published online 24 April) report the results of a meta-analysis and modeling that show that increasing the concentration of atmospheric CO_2 also stimulates microbial decomposition of organic carbon in soils, by roughly the same amount that it increases soil organic carbon, leading to lower equilibrium soil carbon inventories and limiting the accumulation of carbon. Thus, soils may not provide as much carbon storage as hoped.

Abstract

Increasing Turnover
Abstract
Supplementary Material
References and Notes
eLetters (0)

Supplementary Mate

Summary

Materials and Methods

Figs. S1 to S3

Tables S1 to S4

Databases S1 to S6

References (30–106)

Resources

File (1249534s1-s6.xlsx)

DOWNLOAD 236.08 KB

File (vangroenigen.sm.pdf)

DOWNLOAD 1.56 MB

- **Let's get started**

Addendum - order of computations

- **It is important to think about the order in which you apply the statistical computations:**
- effect of increased CO2 on SOC:
  - mean(SOC$_{increased}$) / mean(SOC$_{ambient}$)
  - mean(SOC$_{increased}$ / SOC$_{ambient}$)
- same for log(mean(...)) vs mean(log(...))


Science
Current Issue First release papers Archive About
HOME > SCIENCE > VOL. 344, NO. 6183 > FASTER DECOMPOSITION UNDER INCREASED ATMOSPHERIC CO_2 LIMITS SOIL CARBON STORAGE

REPORT

Faster Decomposition Under Increased Atmospheric CO_2 Limits Soil Carbon Storage

KEES JAN VAN GROENIGEN , XUAN QI, CRAIG W. OSENBERG, YIQI LUO, AND BRUCE A. HUNGATE Authors Info & Affiliations

SCIENCE • 24 Apr 2014 • Vol 344, Issue 6183 • pp. 508-509 • DOI: 10.1126/science.1249534

2,522 272 CHECK ACCESS

Increasing Turnover
Abstract
Supplementary Material
References and Notes
eLetters (0)

Increasing Turnover

Rising concentrations of CO_2 in the atmosphere are likely to stimulate plant growth, leading to increased biomass and soil carbon stocks, thereby lessening the rate of the rise in atmospheric CO_2 levels. **Van Groenigen *et al.*** (p. 508, published online 24 April) report the results of a meta-analysis and modeling that show that increasing the concentration of atmospheric CO_2 also stimulates microbial decomposition of organic carbon in soils, by roughly the same amount that it increases soil organic carbon, leading to lower equilibrium soil carbon inventories and limiting the accumulation of carbon. Thus, soils may not provide as much carbon storage as hoped.

Abstract