

Projekt

Maximilian Ernst

4 1 2021

Datensatz

Der vorliegende Datensatz stammt aus der Studie ‘Kossakowski, J.J., Groot, P.C., Haslbeck, J.M.B., Borsboom, D. and Wichers, M., 2017. Data from ‘Critical Slowing Down as a Personalized Early Warning Signal for Depression’. Journal of Open Psychology Data, 5(1), p.1. DOI: <http://doi.org/10.5334/jopd.29>’.

Es wurden im Zeitraum 2012-2013 zu 1478 Messzeitpunkten Daten einer männlichen, 57 Jahre alten Testperson erhoben. Diese war an einer schweren Depression erkrankt und nahm zu Studienbeginn bereits für 8.5 Jahre Antidepressiva.

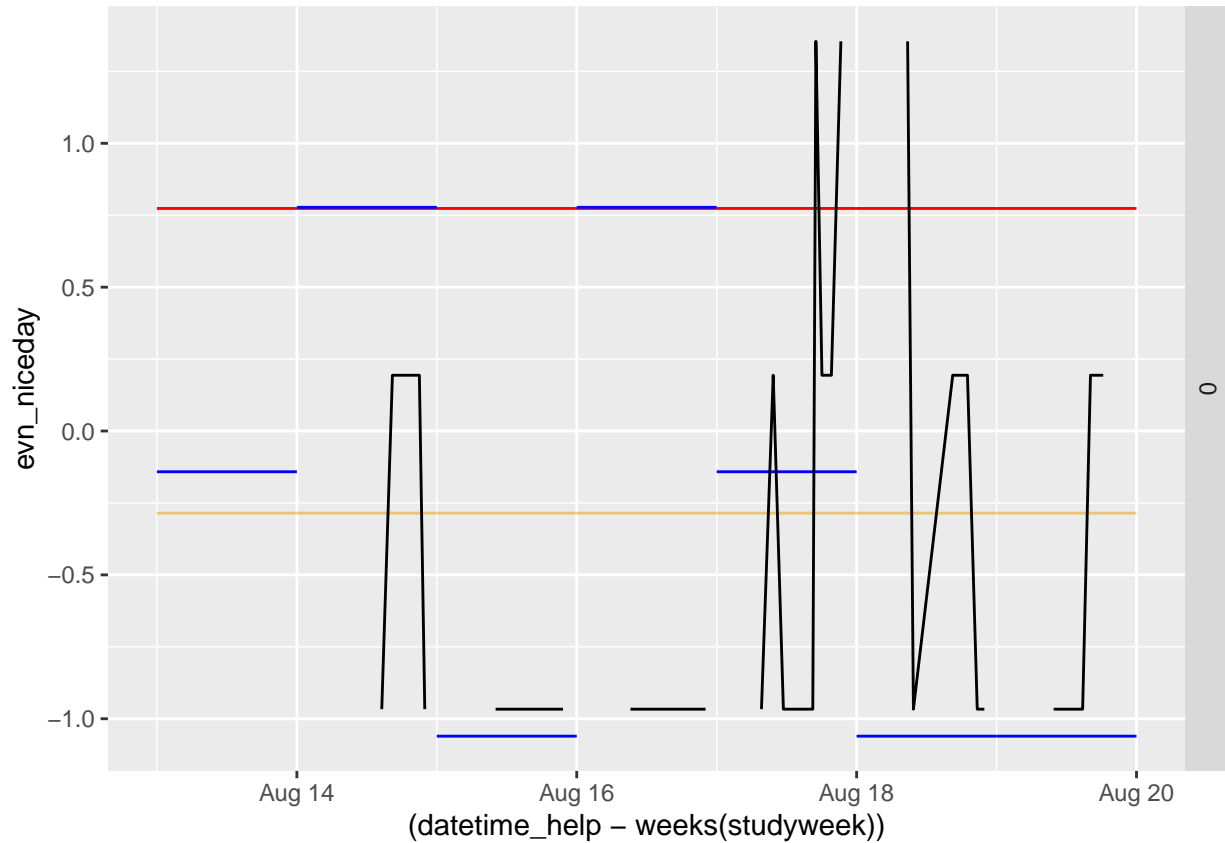
Die für uns relevanten Variablen wurden auf drei verschiedenen Zeitskalen erhoben: Wöchentlich wurde eine Depressionsskala erhoben, Täglich abends das Item “Was this a nice day?” sowie 10 mal täglich Items zur Stimmung, genutzt wurde hier das Item “XXX”. (Es wurden noch viele weitere Items erhoben, die allerdings für die Visualisierung nicht relevant sind).

Des weiter fand eine Intervention statt; die Antidepressiva wurden über den Studienverlauf graduell von 150mg auf 0mg reduziert. Dabei wussten weder die Versuchsperson noch die Forscher, zu welchem Zeitpunkt genau die Reduktion beginnen wird. Allerdings wusste die VP den spätmöglichen Zeitpunkt, zu welchem die Antidepressiva auf 0 reduziert sein müssen.

Darstellung 1: Überblick

Aufgrund der sehr komplexen Struktur der Daten habe ich mich zunächst dazu entschieden, eine Visualisierung zu erstellen, die ein einfaches Verständnis der Datenstruktur ermöglicht.

Dazu habe ich zunächst die Messungen der Items verschiedener Zeitskalen in einer Visualisierung dargestellt:

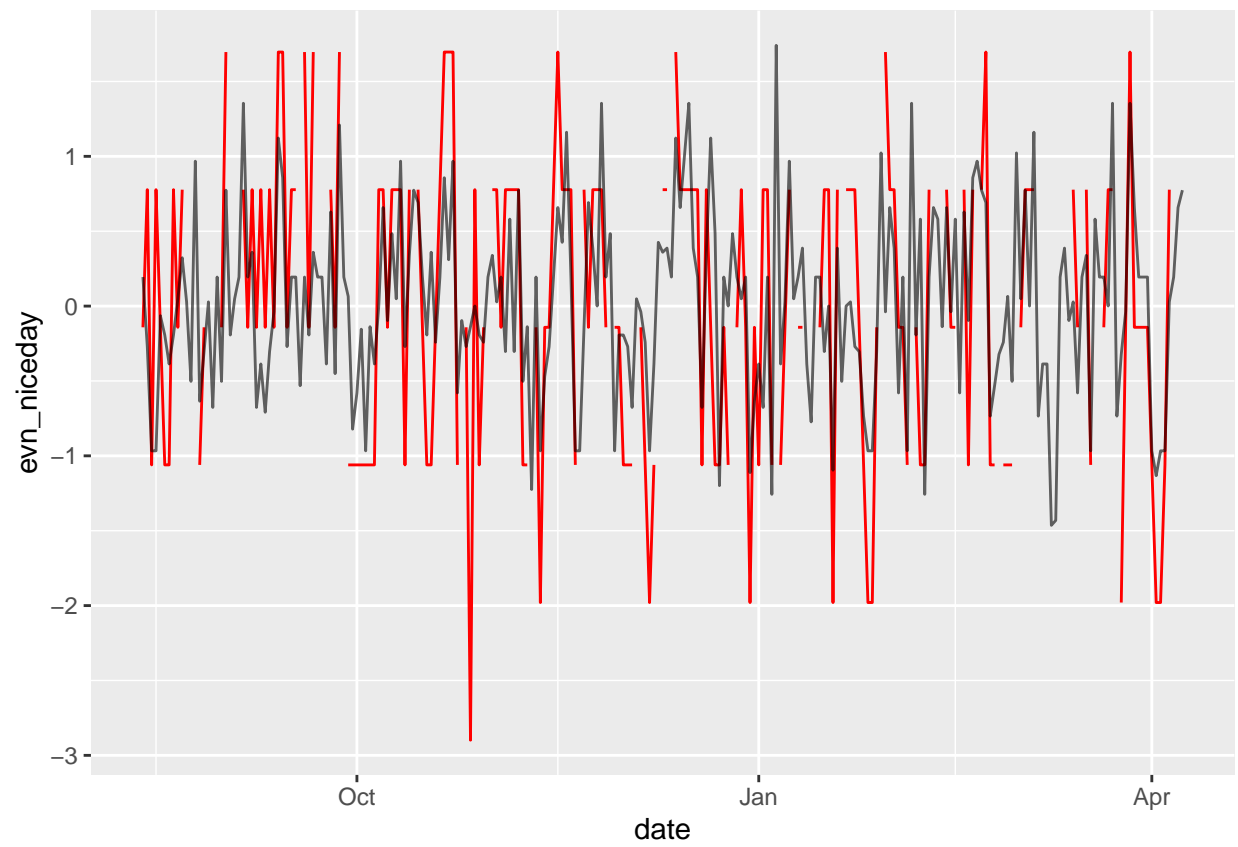


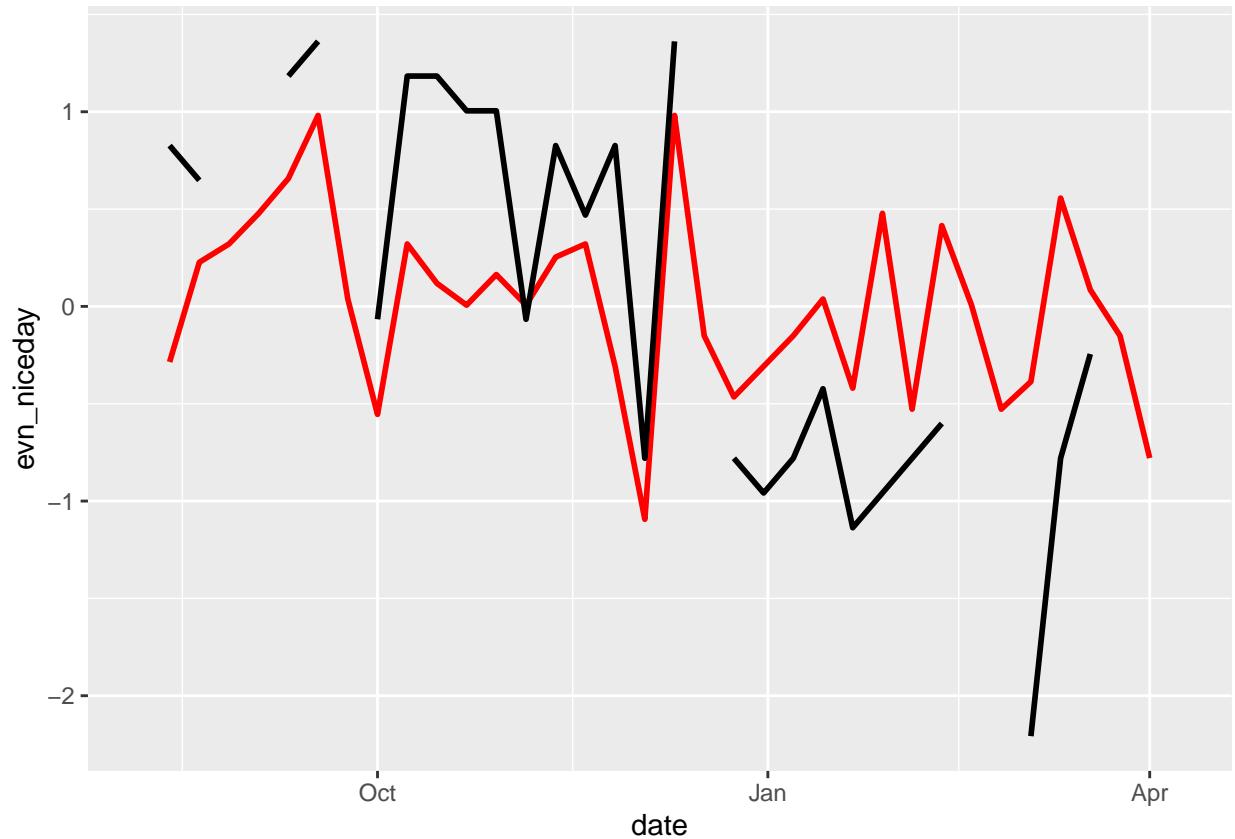
Darstellung 2: Effekt

Zum zweiten möchte ich einen Vergleich zwischen den erhobenen Items auf verschiedenen Zeitskalen anstellen. Dazu habe ich 1. Aus dem mehrmals täglich erhobenen Item den Tagesmittelwert gebildet - um ihn mit dem täglich erhobenen Item zu vergleichen 2. Aus dem täglich erhobenen Item den Wochenmittelwert gebildet - um diesen mit dem wöchentlich erhobenen Item zu vergleichen.

?noch genauer auf Itemauswahl eingehen?

Für den Vergleich habe ich dann die folgenden beiden Grafiken erstellt:





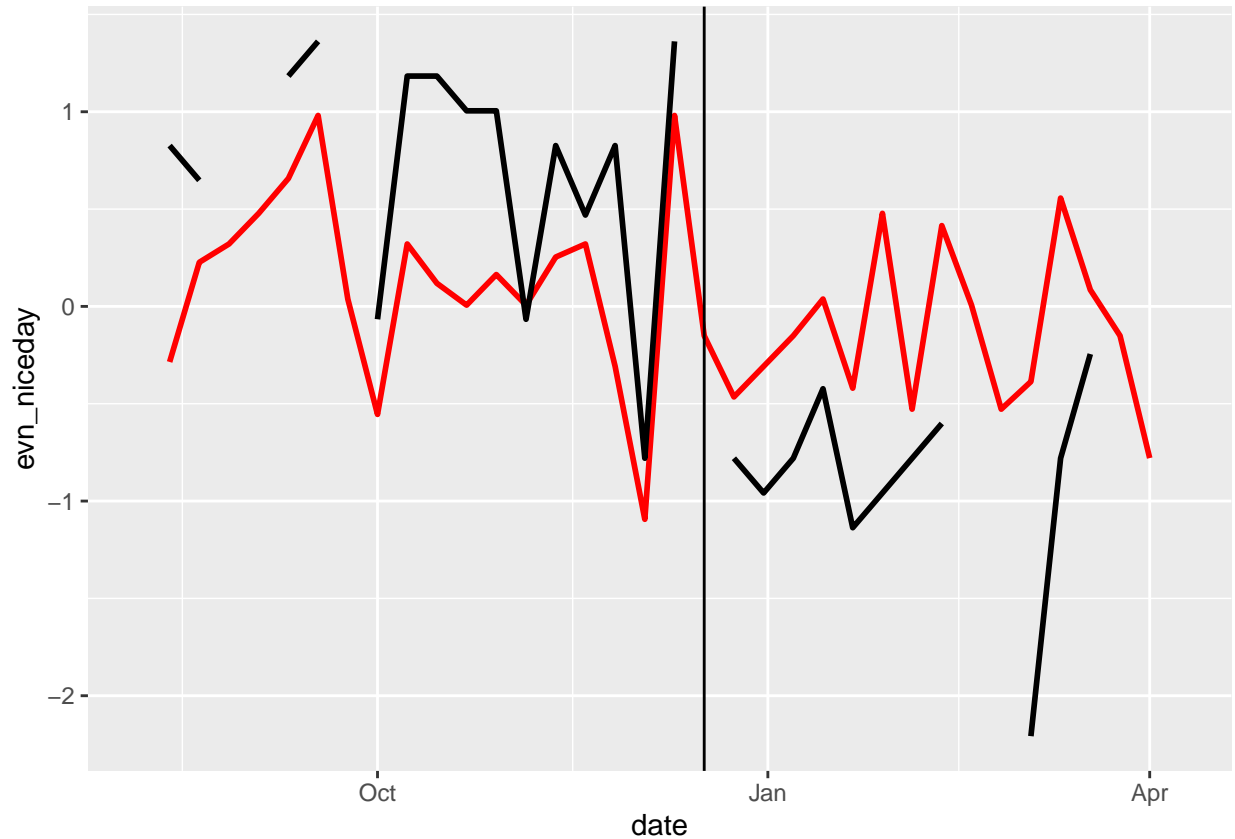
Man sieht, dass es keine systematische Veränderung im ersten Plot über die Zeit gibt, im zweiten Plot sieht man jedoch, dass sie die depressiven Symptome über die Zeit hinweg verschlimmern, während der Wochenmittelwert “schöner Tage” konstant bleibt. Diesem Effekt habe ich mich dann entschieden genauer zu untersuchen.

Dazu habe ich zunächst verschiedene weitere Aggregationsmethoden außer des Wochenmittels genutzt (max, min, sd) um zu überprüfen, ob diese stärker mit der Depressionsskala zusammenhängen, was jedoch nicht der Fall ist.

Dann habe ich den Zeitpunkt eingezeichnet, zu welchem der Testperson klar sein musste, dass sie keine Antidepressiva mehr erhält:

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```



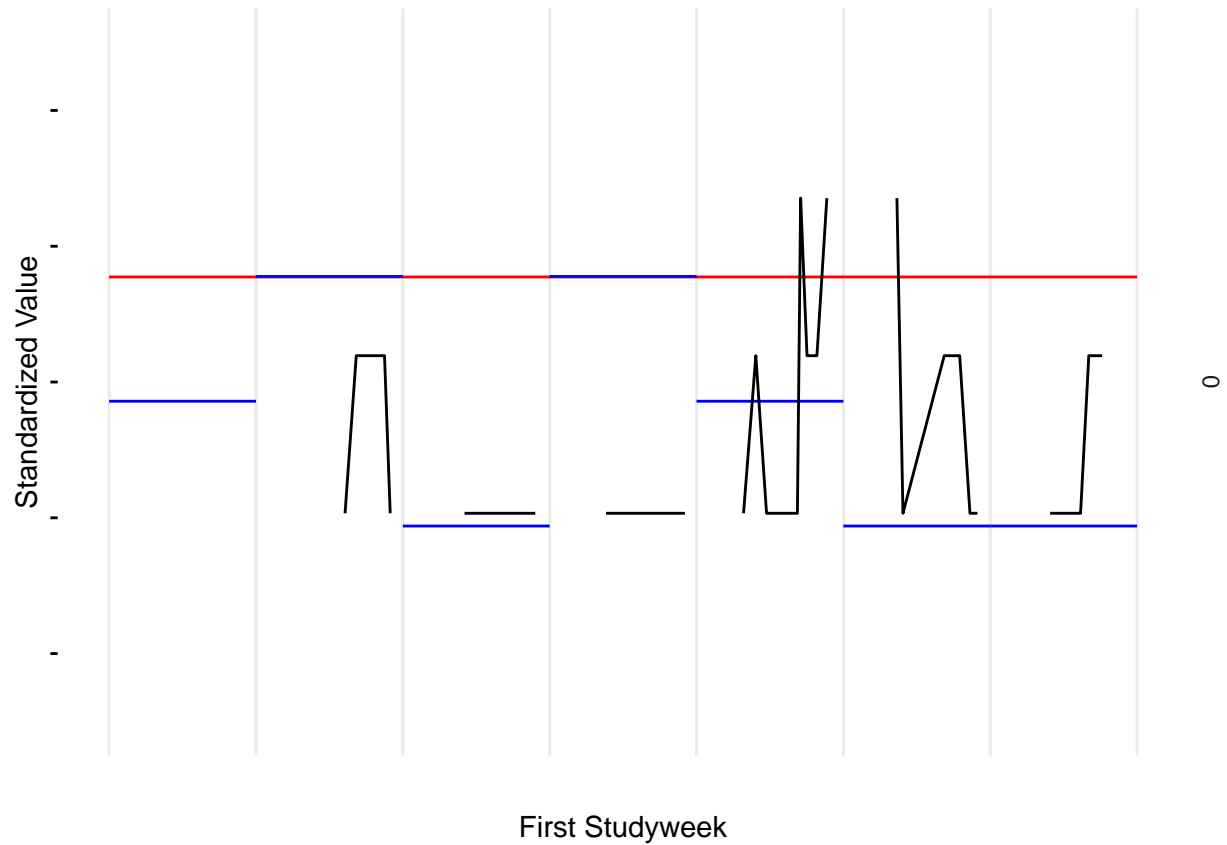
Diese Visualisierung legt eher den Schluss nahe, dass es sich um einen Placebo effekt handelt - das tägliche Item zeigt keine systematische Verschlechterung, die Depressionsskala schon.

Verbesserung der Grafiken

Zunächst habe ich mich dazu entschieden, in der Überblicksgrafik nur Rohdaten zu zeigen.

Data Ink Ratio

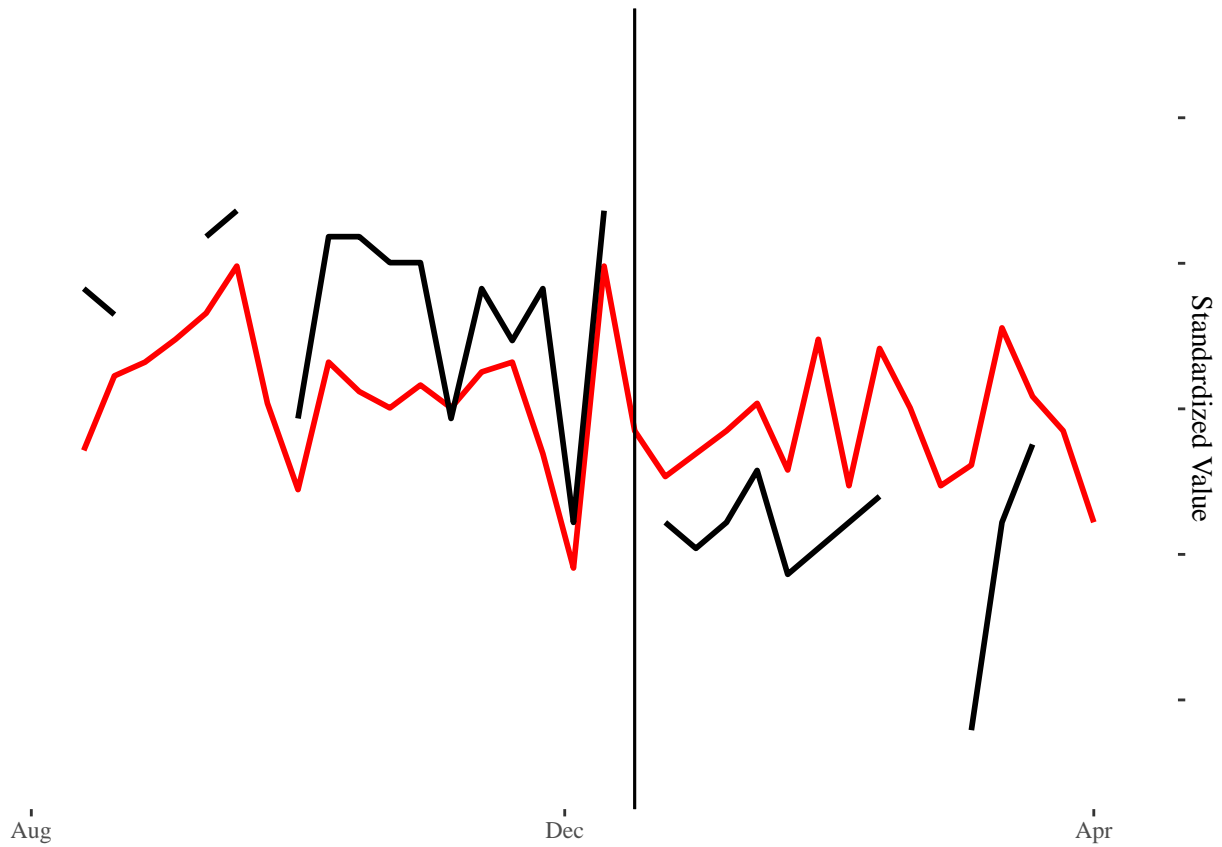
Dann habe ich die Data-Ink Ratio der Grafiken verbessert:



I decided to set vertical lines at the end of each day to resemble the familiar look of a calendar. I also decided to set y-Axis breaks at the first and second standard deviation. (I have no idea where the 0 on the right side of the plot comes from; I later removed that in Inkscape)

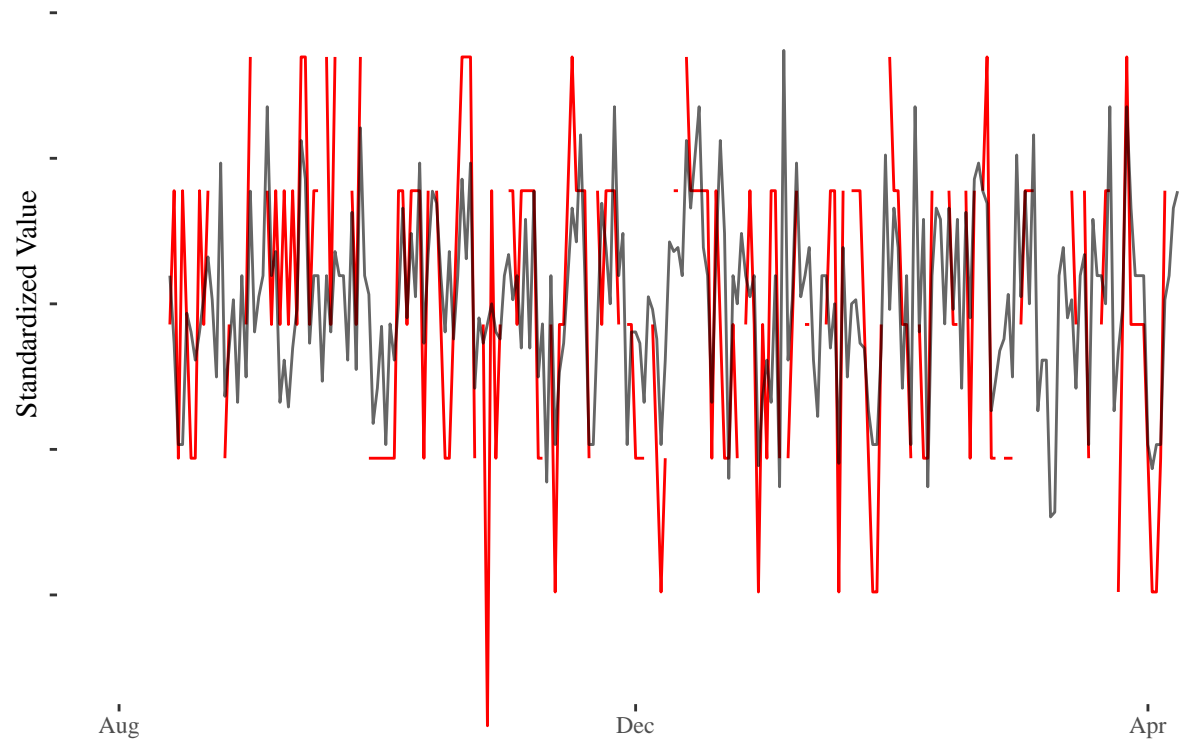
```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```



The y axis label is moved to the right because of how the plots are put together later.

`## Warning: Removed 1 row(s) containing missing values (geom_path).`



Farben und Schrift

Als nächstes habe ich dann eine Farbskala ausgesucht, die robust gegenüber Farbenblindheit ist. Es handelt sich um die Skala “<https://rdr.io/cran/ggthemes/man/colorblind.html>”:

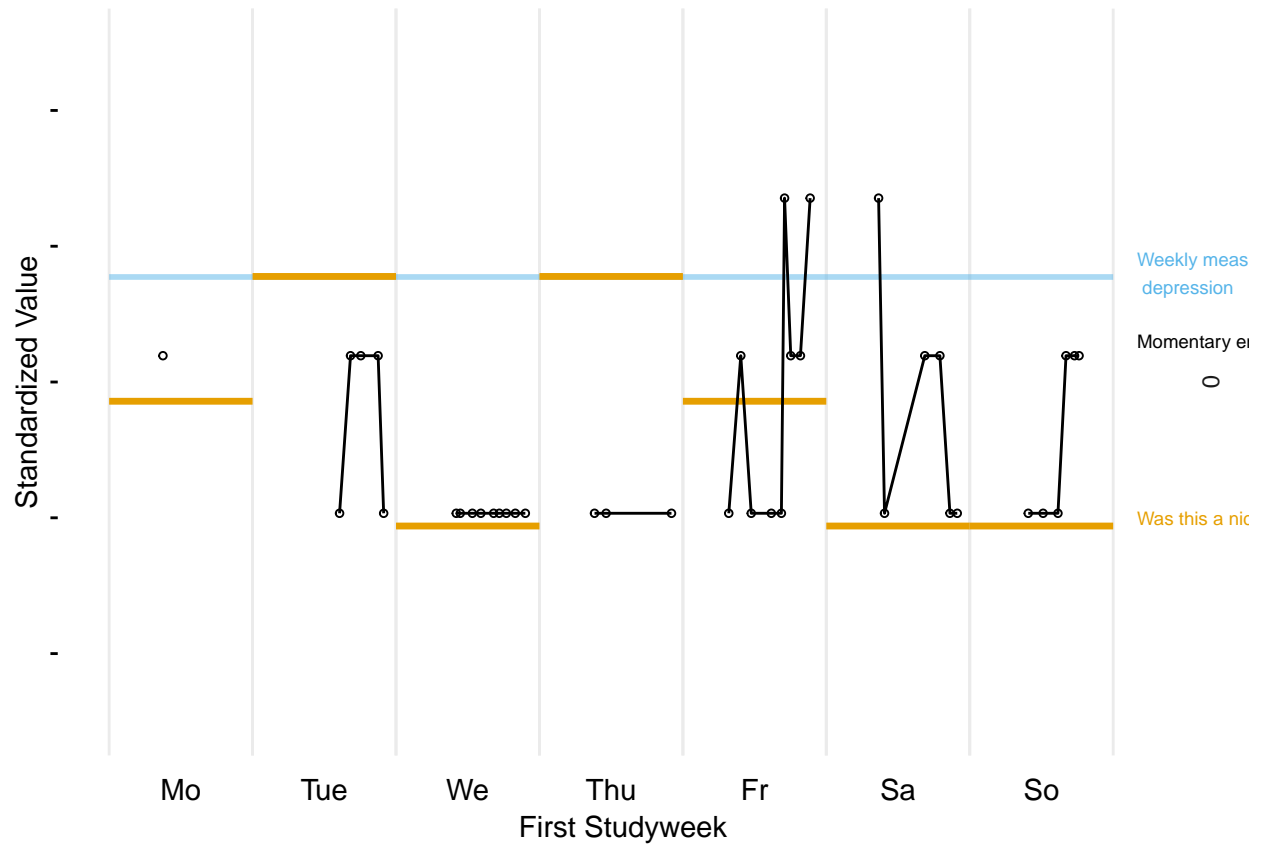


Allerdings ist der Helligkeitsunterschied zwischen Blau und Gelb zu gering, weshalb sie in schwarz-weiß schlecht unterscheidbar sind:

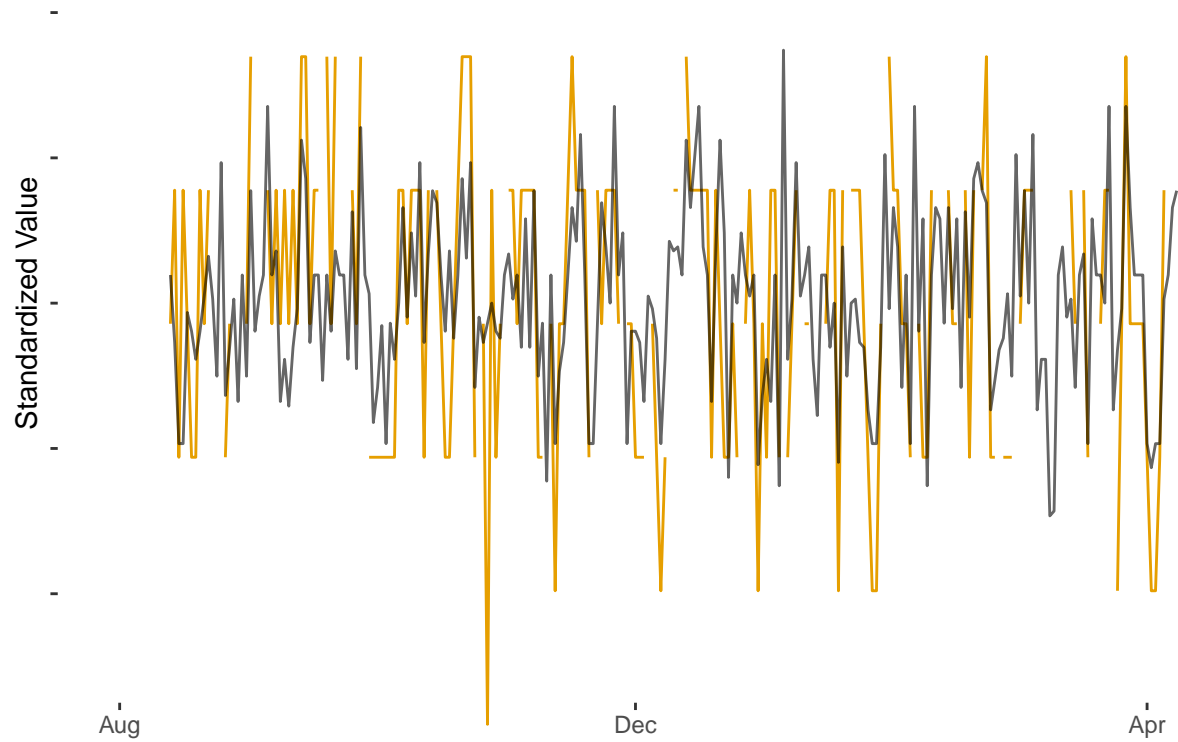
Ich habe mich deshalb dazu entschlossen, den alpha-Wert des Blautons etwas herunterzusetzen.

```
## Warning: Removed 31 rows containing missing values (geom_point).
```

```
## Warning: Removed 31 row(s) containing missing values (geom_path).
```

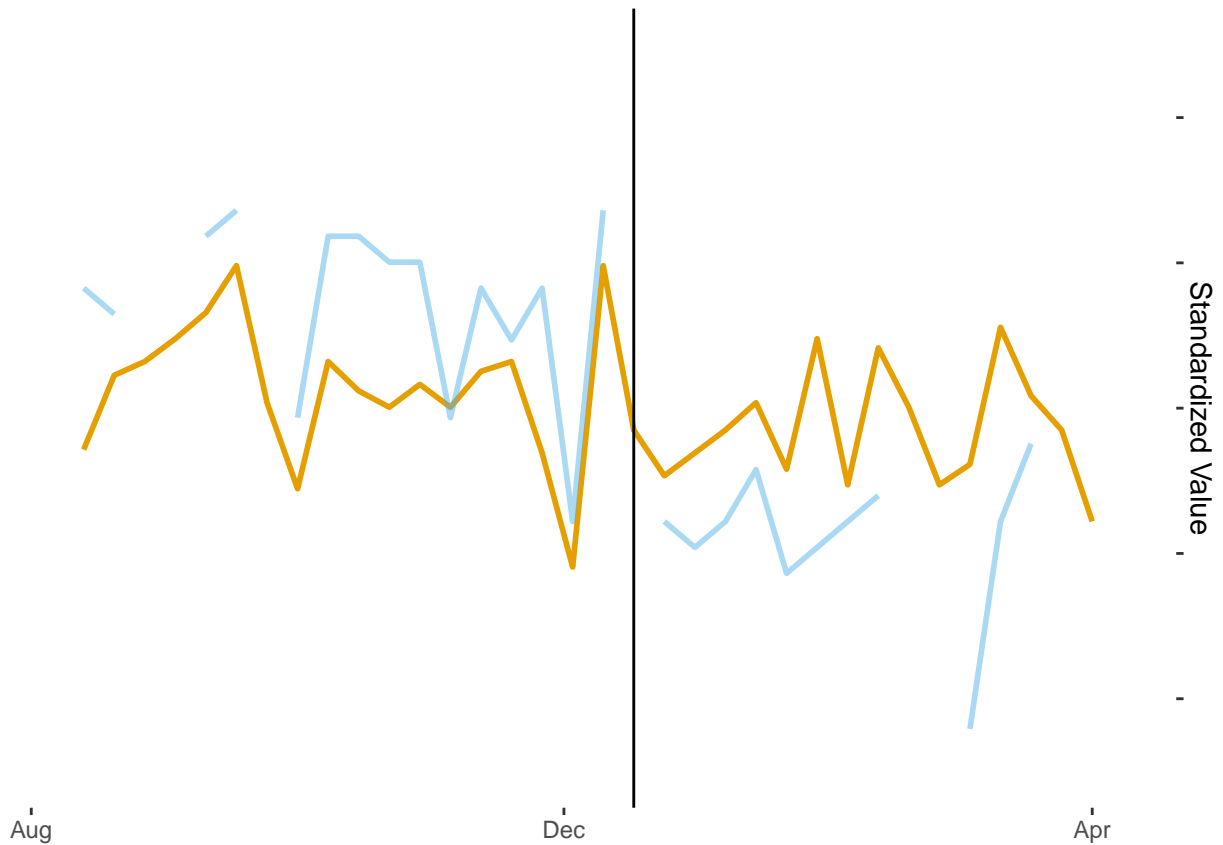


Warning: Removed 1 row(s) containing missing values (geom_path).



```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```

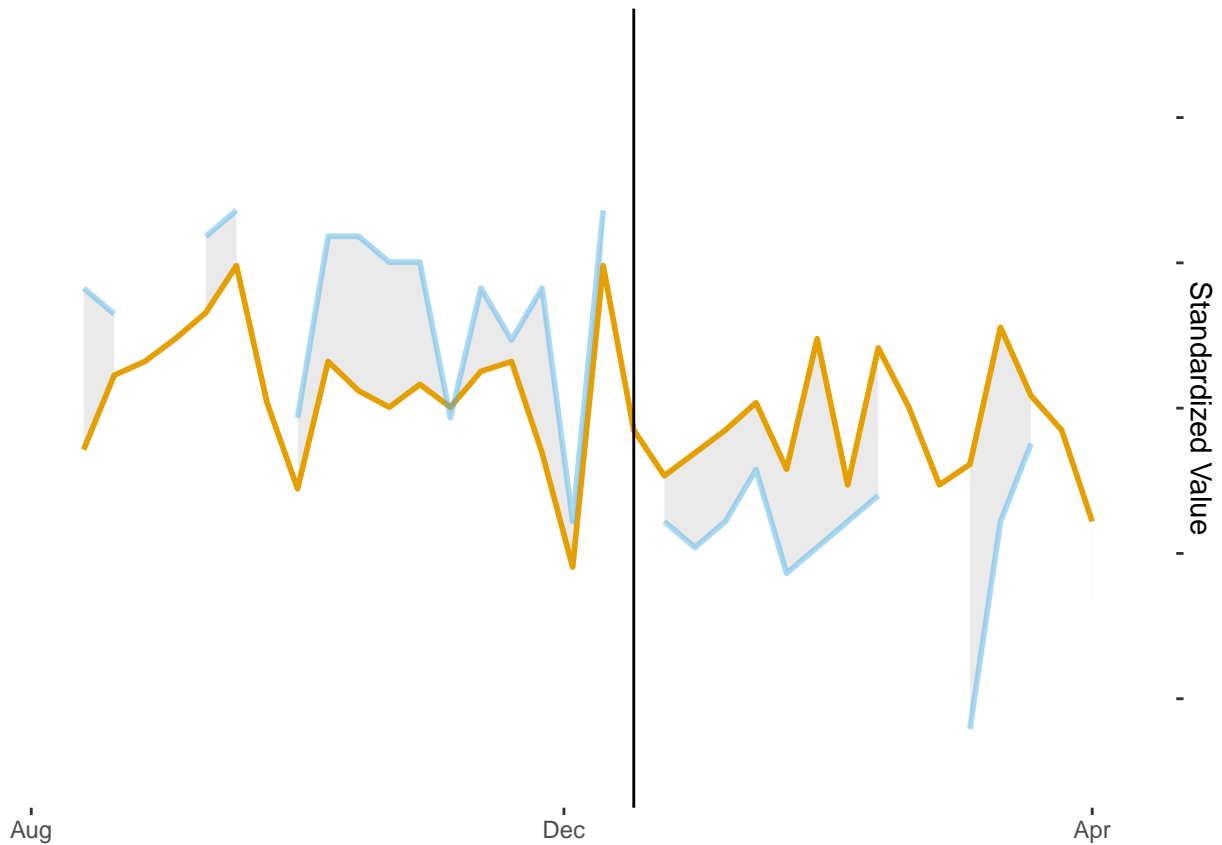


Weitere Verbesserungen

Außerdem habe ich die Fläche im wöchentlichen Plot eingefärbt, da vor allem der unterschied zwischen täglicher und wöchentlicher Messung von Belang ist:

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```



Ich habe den Flächen ein Vorzeichen gegeben, fand es damit allerdings eher verwirrend:

Layout

Der Betrachter sollte zunächst die Überblicksgrafik anschauen, um sich ein Bild von den Daten zu machen, weshalb ich diese auf der linken Seite platziert habe. Die wöchentliche Grafik habe ich dann rechts davon platziert, da die y - Skala gleich ist und somit der Zusammenhang zwischen den Grafiken besser ersichtlich wird. Um diesen weiter zu verdeutlichen, habe ich die korrespondierenden Daten gleich eingefärbt und mit Pfeilen verbunden, die mit der Aggregationsmethode beschriftet sind. Da die Grafik mit den täglich erhobenen Daten sehr viele Messzeitpunkte enthält und deshalb schwer zu lesen ist, und außerdem das Layout verkompliziert, habe ich mich dazu entschieden sie wegzulassen.