

# Individual assignment SBD

Hagen Aad Fock

ADS, 2021-2022

## Contents

<b>0. Prepare</b>	<b>1</b>
0.1 Checking for missing data . . . . .	2
<b>1. General</b>	<b>12</b>
1.1. Describe your data . . . . .	12
1.2. Visualize your data . . . . .	17
<b>2. Forecasting</b>	<b>21</b>
2.1. SARIMA modeling . . . . .	21
2.2. Dynamic regression . . . . .	22
2.3. Forecasts . . . . .	23
<b>3. Causal Modeling</b>	<b>23</b>
3.2 Analysis . . . . .	23
3.2a Granger Causal analysis . . . . .	23
3.2b Interrupted Time Series analysis . . . . .	24
3.2c Synthetic Control Analysis . . . . .	24
3.3 Conclusion and critical reflection . . . . .	24

## 0. Prepare

► Load the R-packages you will use.

```
library(fpp3)
library(tseries)
library(tidyverse)    # alternatively, this also loads %>%
library(knitr)
library(mice) # for missing data imputation
library(VIM)
library(expsmooth)
```

- Include R-code you used to load (and prepare) the data.

## 0.1 Checking for missing data

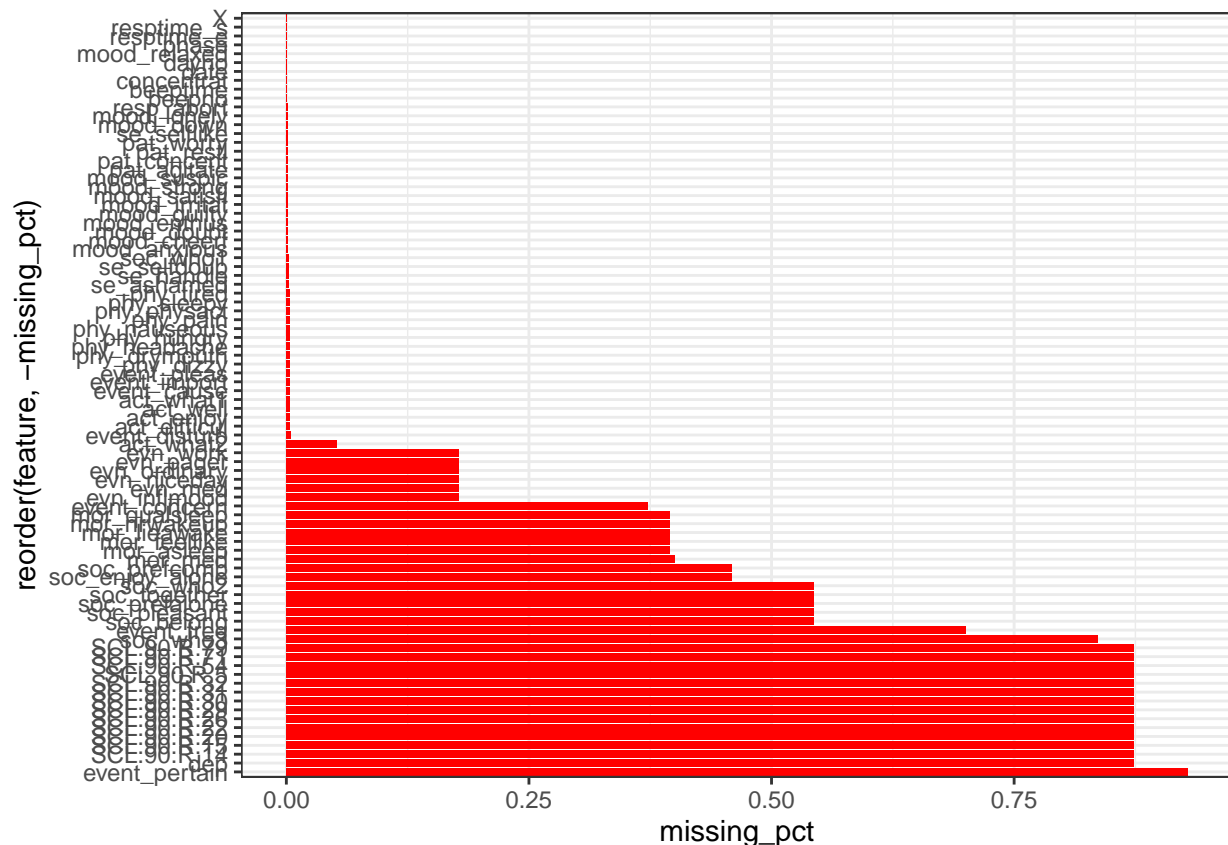
```
set.seed(666)

esmdata <- read.csv("../ESMdata/ESMdata.csv")

missing_values <- esmdata %>%
  summarize_all(funs(sum(is.na(.))/n())) %>%
  gather(key="feature", value="missing_pct")

## Warning: `funs()` was deprecated in dplyr 0.8.0.
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.

missing_values %>%
  ggplot(aes(x=reorder(feature,-missing_pct),y=missing_pct)) +
  geom_bar(stat="identity",fill="red")+
  coord_flip()+theme_bw()
```



As we can see there is a huge part of the data missing. In section 1.1. Describe your data is stated which variables I will use to predict the outcome variable. Luckily I only include variables that have an insignificant percentage of missing data. To impute the missing data I will use the package mice.

```
relevant_variables <- c('date', 'resptime_s', 'concentrat',

                        'mood_relaxed', 'mood_satisfi', 'mood_enthus',
                        'mood_cheerf', 'mood_strong', 'se_selflike',
                        'se_handle',

                        'mood_irritat', 'mood_suspici', 'mood_doubt',
                        'se_ashamed', 'se_selfdoub',

                        'mood_down', 'mood_lonely', 'mood_anxious',
                        'mood_guilty',

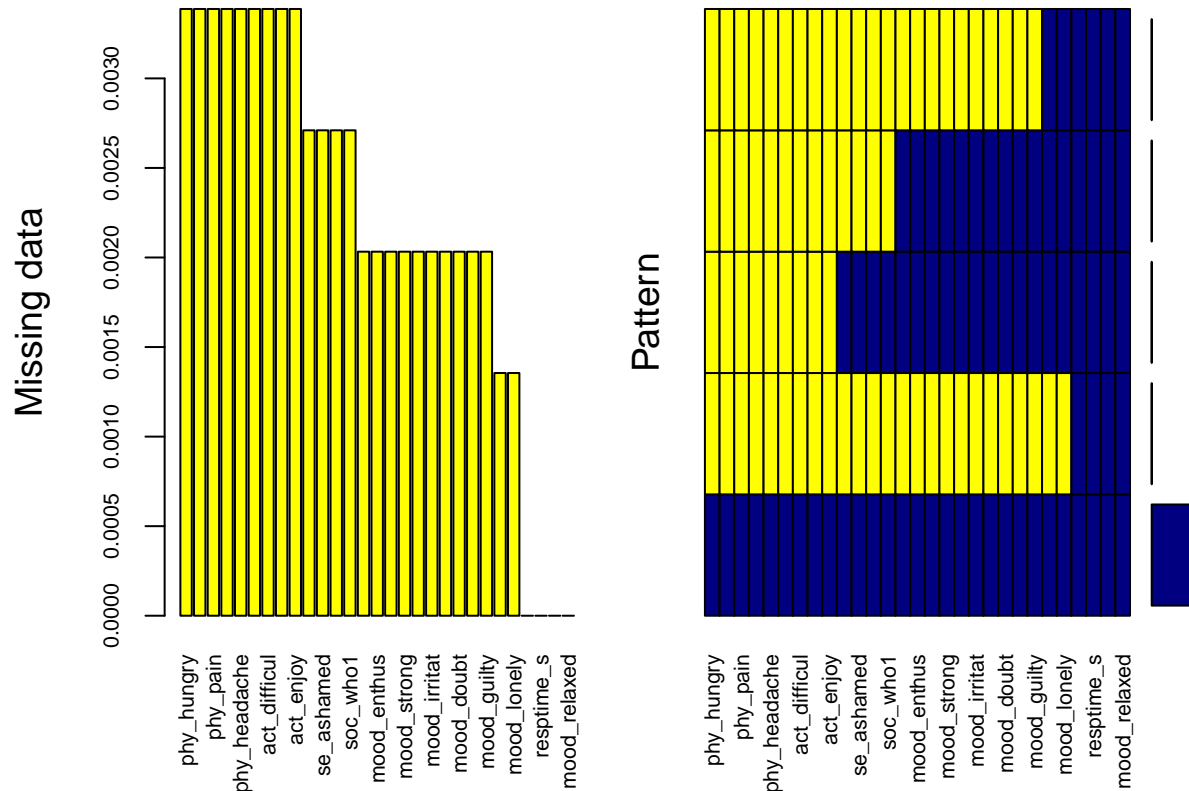
                        'phy_hungry', 'phy_tired', 'phy_pain', 'phy_dizzy',
                        'phy_headache',

                        'soc_who1', 'phy_physact', 'act_difficul',
                        'act_well', 'act_enjoy')
```

```
esmdata <- esmdata[,relevant_variables]

mice_plot <- aggr(esmdata[,relevant_variables], col=c('navyblue','yellow'),
  numbers=TRUE, sortVars=TRUE,
  labels=names(relevant_variables), cex.axis=.7,
  gap=3, ylab=c("Missing data","Pattern"))
```

```
## Warning in plot.aggr(res, ...): not enough horizontal space to display
## frequencies
```



```
##
## Variables sorted by number of missings:
##   Variable      Count
##   phy_hungry 0.003387534
##   phy_tired 0.003387534
##   phy_pain 0.003387534
##   phy_dizzy 0.003387534
##   phy_headache 0.003387534
##   phy_physact 0.003387534
##   act_difficul 0.003387534
##   act_well 0.003387534
##   act_enjoy 0.003387534
##   se_handle 0.002710027
##   se_ashamed 0.002710027
```

```

## se_selfdoub 0.002710027
## soc_who1 0.002710027
## mood_satisfi 0.002032520
## mood_enthus 0.002032520
## mood_cheerf 0.002032520
## mood_strong 0.002032520
## se_selflike 0.002032520
## mood_irritat 0.002032520
## mood_suspici 0.002032520
## mood_doubt 0.002032520
## mood_anxious 0.002032520
## mood_guilty 0.002032520
## mood_down 0.001355014
## mood_lonely 0.001355014
## date 0.000000000
## resptime_s 0.000000000
## concentrat 0.000000000
## mood_relaxed 0.000000000

```

```
##
```

```
## iter imp variable
```

##	1	1	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	1	2	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	1	3	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	1	4	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	1	5	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	2	1	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	2	2	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	2	3	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	2	4	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	2	5	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	3	1	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	3	2	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	3	3	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	3	4	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	3	5	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	4	1	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	4	2	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	4	3	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	4	4	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	4	5	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	5	1	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	5	2	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	5	3	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle
##	5	4	mood_satisfi	mood_enthus	mood_cheerf	mood_strong	se_selflike	se_handle

[illegible]

[illegible]

8

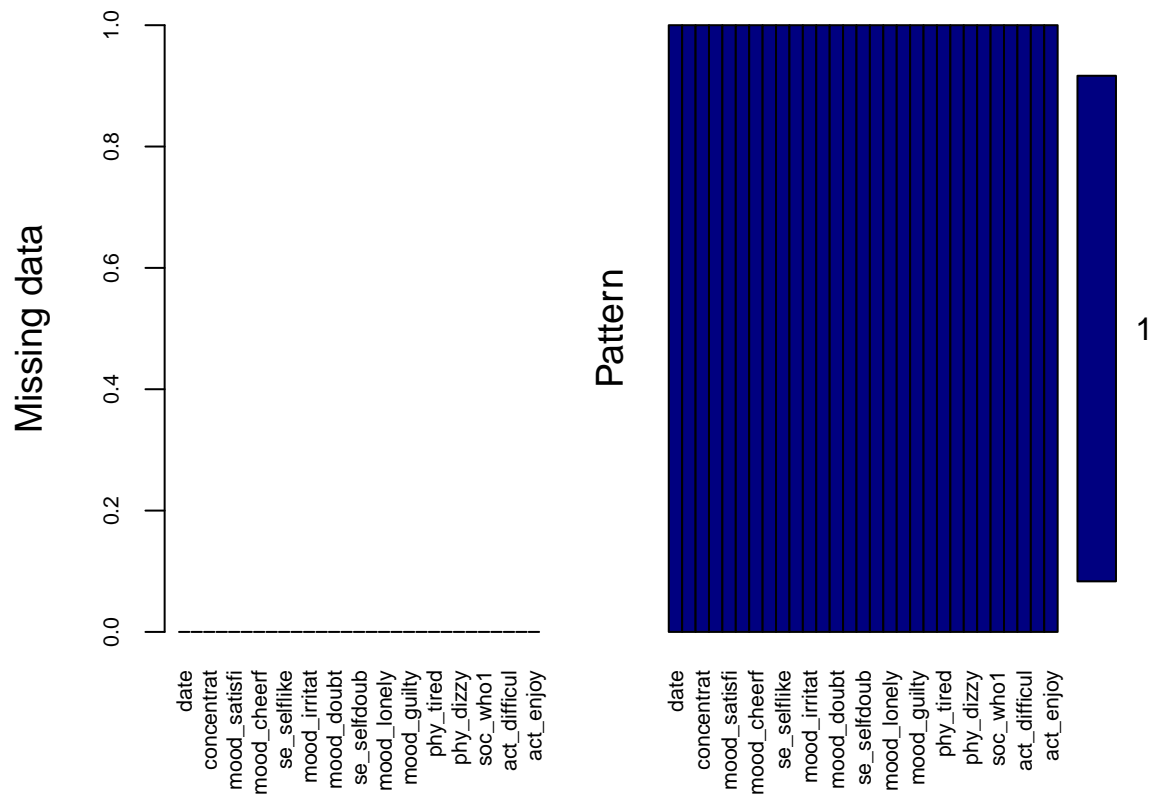


[illegible]

[illegible]

```
## 50 5 mood_satisfi mood_enthus mood_cheerf mood_strong se_selflike se_handle
## Warning: Number of logged events: 2
```

```
mice_plot <- aggr(esmdata, col=c('navyblue','yellow'),
  numbers=TRUE, sortVars=TRUE,
  labels=names(esmdata), cex.axis=.7,
  gap=3, ylab=c("Missing data","Pattern"))
```



```
##
## Variables sorted by number of missings:
## Variable Count
## date 0
## resptime_s 0
## concentrat 0
## mood_relaxed 0
## mood_satisfi 0
## mood_enthus 0
## mood_cheerf 0
## mood_strong 0
## se_selflike 0
## se_handle 0
## mood_irritat 0
## mood_suspici 0
## mood_doubt 0
```

```
##      se_ashamed      0
##      se_selfdoub     0
##      mood_down      0
##      mood_lonely     0
##      mood_anxious    0
##      mood_guilty     0
##      phy_hungry      0
##      phy_tired       0
##      phy_pain        0
##      phy_dizzy       0
##      phy_headache    0
##      soc_who1        0
##      phy_physact     0
##      act_difficul    0
##      act_well        0
##      act_enjoy       0

esmdata$date <- dmy(esmdata$date)
esmdata <- aggregate(esmdata[, 3:29], list(esmdata$date), mean)
names(esmdata)[1] <- 'date'
```

Now that every missing data was imputed we can make the tsibble.

```
create_date_time <- function(date, time) {
  return(dmy_hms(paste(date, time, sep=" ")))
}

transform_soc_who1 <- function(soc_who1) {
  return (if (soc_who1 > 0) 1 else 0)
}

esmdata$soc_who1 <- sapply(esmdata$soc_who1, transform_soc_who1)

esmdata <- esmdata %>%
  as_tsibble(index = date)
```

## 1. General

► To be able to use fpp3, the data have to be a tsibble object. If they aren't already, transform them. Describe the structure of this object.

### 1.1. Describe your data

The data is about a man that was reducing his anti depression medication. Every variable that will be stated within this section was measured using a semi-random experience-

sampling protocol. “The participant collected reports of momentary states up to 10 times a day over a period of 239 days.” <https://doi.org/10.1159/000441458>

“Depression is a mood disorder that causes a persistent feeling of sadness and loss of interest. Also called major depressive disorder or clinical depression, it affects how you feel, think and behave and can lead to a variety of emotional and physical problems. You may have trouble doing normal day-to-day activities, and sometimes you may feel as if life isn’t worth living.” <https://www.mayoclinic.org/diseases-conditions/depression/symptoms-causes/syc-20356007>

► What is your outcome variable; how was it measured (how many times, how frequently, etc.)?

Regarding the information from the article about depression I derived following variables as potential outcome variables:

I don’t have many experience regarding depression but I assume that depression is simplified a bad mood over a longer period. Therefore, I derived all mood variables and came up with a mixed mood variable. I call it the depression factor.

Because the scales are not always displayed by the same scale I had to normalize them. I transformed all scales to the scheme (-100, 100) so minus values indicate a bad influence for the mood and positive values indicate good moods. Also I wanted to make the range bigger so the changes are more visible.

name	description	scale	~ missing %
mood_relaxed	I feel relaxed	positive ( 1, 7)	0%
mood_satisfi	I feel satisfied	positive ( 1, 7)	0.2%
mood_enthus	I feel enthusiastic	positive ( 1, 7)	0.2%
mood_cheerf	I feel cheerful	positive ( 1, 7)	0.2%
mood_strong	I feel strong	positive ( 1, 7)	0.2%
se_selflike	I like myself	positive ( 1, 7)	0.2%
se_handle	I can handle anything	positive ( 1, 7)	0.3%
mood_irritat	I feel irritated	negative ( 1, 7)	0.2%
mood_suspici	I feel suspicious	negative ( 1, 7)	0.2%
mood_doubt	I feel indecisive	negative ( 1, 7)	0.2%
se_ashamed	I am ashamed of myself	negative ( 1, 7)	0.3%
se_selfdoub	I doubt myself	negative ( 1, 7)	0.3%
mood_down	I feel down	negative (-3, 3)	0.1%
mood_lonely	I feel lonely	negative (-3, 3)	0.1%
mood_anxious	I feel anxious	negative (-3, 3)	0.2%
mood_guilty	I feel guilty	negative (-3, 3)	0.2%

► What are the predictor variable(s) you will consider? Why would this make sense as a predictor?

I assume that physical condition, and the previous day might be predictor. For example when the patient is in pain he might feel more depressed and or the previous day was an nice day he is more motivated about the upcoming day.

name	description	scale	~ missing %
phy_hungry	I am hungry	negative ( 1, 7)	0.3%
phy_tired	I am tired	negative ( 1, 7)	0.3%
phy_pain	I am in pain	negative ( 1, 7)	0.3%
phy_dizzy	I feel dizzy	negative ( 1, 7)	0.3%
phy_headache	I have a headache	negative ( 1, 7)	0.3%

► What are the cause(s) you will consider? Why would this make sense as a cause?

I assume that a human being is complex but an encapsulated system on its own. But the surrounding factors influence whatever happens within a person. So I assume all activities with the environment like meeting people or doing something actively must be the cause of how the human being feels and also everything a person takes to him influences the internal behavior.

name	description	scale	~ missing %
soc_who1	Who am I with?	( 1, 7)	3%
soc_enjoy_alone	I enjoy to be alone	( 1, 7)	46%
soc_prefcomp	I prefer being in company	( 1, 7)	46%
soc_pleasant	I find this company pleasant	( 1, 7)	54%
soc_prefalone	I prefer to be alone	( 1, 7)	46%
phy_physact	From the last beep on wards I was physically active	( 1, 7)	3.4%

name	description	scale	~ missing %
act_what1	What am I doing (right before the beep)?	0 = nothing 1 = resting 10 = work/studies 20 = housekeeping/shopping 21 = caring for others 26 = medical care 27 = taking care of oneself 41 = sports 43 = active relaxation 45 = passive relaxation 47 = chat-ting/texting/facebook etc 49 = chilling 51 = talking 60 = eating/drinking 88 = traveling 89 = other	3.4%
act_difficul	This (activity) requires effort	( 1, 7)	3.4%
act_well	I am good at this	( 1, 7)	3.4%
act_enjoy	I like doing this	( 1, 7)	3.4%

```

# the variable declaration is following
# _positive indicating if the highest value is a good mood
# _negative indicating if the highest value is a depressive mood
# _17      indicating if the scale is ( 1, 7)
# _33      indicating if the scale is (-3, 3)

# Data Normalization - Min-Max Normalization
# from (-3, 3) to (-100, 100)
normalize <- function(atr, min, max){
  return((atr - (min)) / (max - (min)) * (100 - (-100)) + (-100))
}

# For attributes on a scale (1, 7) where 1 indicates a depressive value.
normalize_positive_17 <- function(atr) (
  return(normalize((atr - 4), -3, 3))
)

normalize_negative_17 <- function(atr) (

```

```

    return(normalize(((atr - 4) * -1), -3, 3))
)

# For attributes on a scale(-3, 3) where 3 is a depressive value.
normalize_negative_33 <- function(atr) (
  return(normalize((atr * -1), -3, 3))
)

# mood_cat1 - described as positive moods
esmdata$mood_relaxed <- normalize_positive_17(esmdata$mood_relaxed)
esmdata$mood_satisfi <- normalize_positive_17(esmdata$mood_satisfi)
esmdata$mood_enthus <- normalize_positive_17(esmdata$mood_enthus)
esmdata$mood_cheerf <- normalize_positive_17(esmdata$mood_cheerf)
esmdata$mood_strong <- normalize_positive_17(esmdata$mood_strong)
esmdata$sse_selflike <- normalize_positive_17(esmdata$sse_selflike)
esmdata$sse_handle <- normalize_positive_17(esmdata$sse_handle)

# mood_cat2 - described as negative moods
esmdata$mood_irritat <- normalize_negative_17(esmdata$mood_irritat)
esmdata$mood_suspici <- normalize_negative_17(esmdata$mood_suspici)
esmdata$mood_doubt <- normalize_negative_17(esmdata$mood_doubt)
esmdata$sse_ashamed <- normalize_negative_17(esmdata$sse_ashamed)
esmdata$sse_selfdoub <- normalize_negative_17(esmdata$sse_selfdoub)

# mood_cat3 - described as depressive moods
esmdata$mood_down <- normalize_negative_33(esmdata$mood_down)
esmdata$mood_lonely <- normalize_negative_33(esmdata$mood_lonely)
esmdata$mood_anxious <- normalize_negative_33(esmdata$mood_anxious)
esmdata$mood_guilty <- normalize_negative_33(esmdata$mood_guilty)

# physical - described as all physical attributes
esmdata$phy_hungry <- normalize_negative_17(esmdata$phy_hungry)
esmdata$phy_tired <- normalize_negative_17(esmdata$phy_tired)
esmdata$phy_pain <- normalize_negative_17(esmdata$phy_pain)
esmdata$phy_dizzy <- normalize_negative_17(esmdata$phy_dizzy)
esmdata$phy_headache <- normalize_negative_17(esmdata$phy_headache)

esmdata$positive_moods <- esmdata$mood_relaxed + esmdata$mood_satisfi +
  esmdata$mood_enthus + esmdata$mood_cheerf + esmdata$mood_strong +
  esmdata$sse_selflike + esmdata$sse_handle

esmdata$negative_moods <- esmdata$mood_irritat + esmdata$mood_suspici +
  esmdata$mood_doubt + esmdata$sse_ashamed + esmdata$sse_selfdoub

esmdata$depressive_moods <- esmdata$mood_down + esmdata$mood_lonely +

```



```

esmdata$mood_anxious + esmdata$mood_guilty

esmdata$depression_factor <- esmdata$positive_moods + esmdata$negative_moods +
  esmdata$depressive_moods

esmdata$physical <- esmdata$phy_hungry + esmdata$phy_tired + esmdata$phy_pain +
  esmdata$phy_dizzy + esmdata$phy_headache

esmdata$positive_moods <- normalize(esmdata$positive_moods,
                                   min(esmdata$positive_moods),
                                   max(esmdata$positive_moods))

esmdata$negative_moods <- normalize(esmdata$negative_moods,
                                   min(esmdata$negative_moods),
                                   max(esmdata$negative_moods))

esmdata$depressive_moods <- normalize(esmdata$depressive_moods,
                                     min(esmdata$depressive_moods),
                                     max(esmdata$depressive_moods))

esmdata$depression_factor <- normalize(esmdata$depression_factor,
                                       min(esmdata$depression_factor),
                                       max(esmdata$depression_factor))

esmdata$physical <- normalize(esmdata$physical,
                             min(esmdata$physical),
                             max(esmdata$physical))

esmdata <- na.omit(esmdata)

```

## 1.2. Visualize your data

► Create a sequence plot of the data with the function `autoplot()`. Interpret the results.

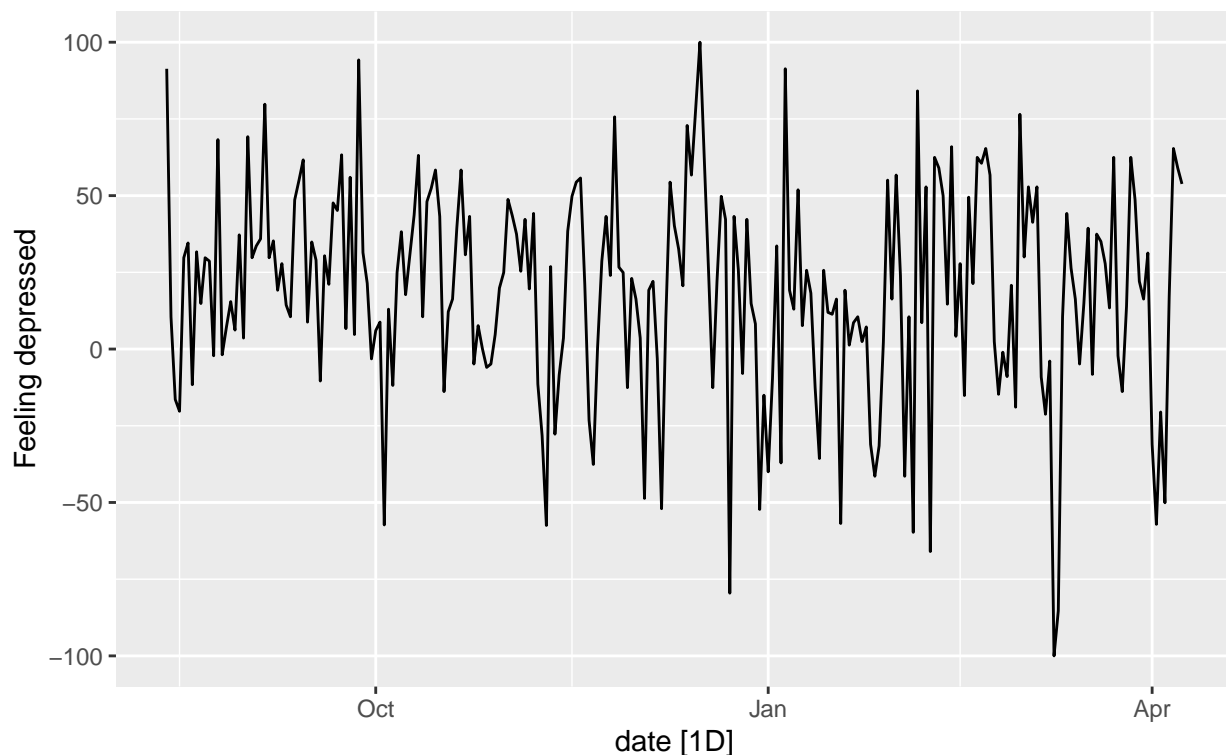
```

autoplot(esmdata, depression_factor) +
  labs(title = "Accumulated Depressive Moods",
       subtitle = "Negative values = bad mood; Positive values = good mood",
       y = "Feeling depressed")

```

## Accumulated Depressive Moods

Negative values = bad mood; Positive values = good mood



- We can see that there are by far more “felling down” samples then “feeling uplifted” samples.
- We can also recognize that the second section of the time series has more extreme values then the first section.
- There are only three samples where the patent was extremely uplifted.
- It seems like the patent had balanced feelings in the first section and one-sided feelings towards “feeling down” in the second section.
- We can see periods where one emotion is more present. For example the patent felt down just at the beginning of the trial.

► Plot the autocorrelation function with the function `acf()`. Interpret the results.

```
# positive moods
```

```
?acf
```

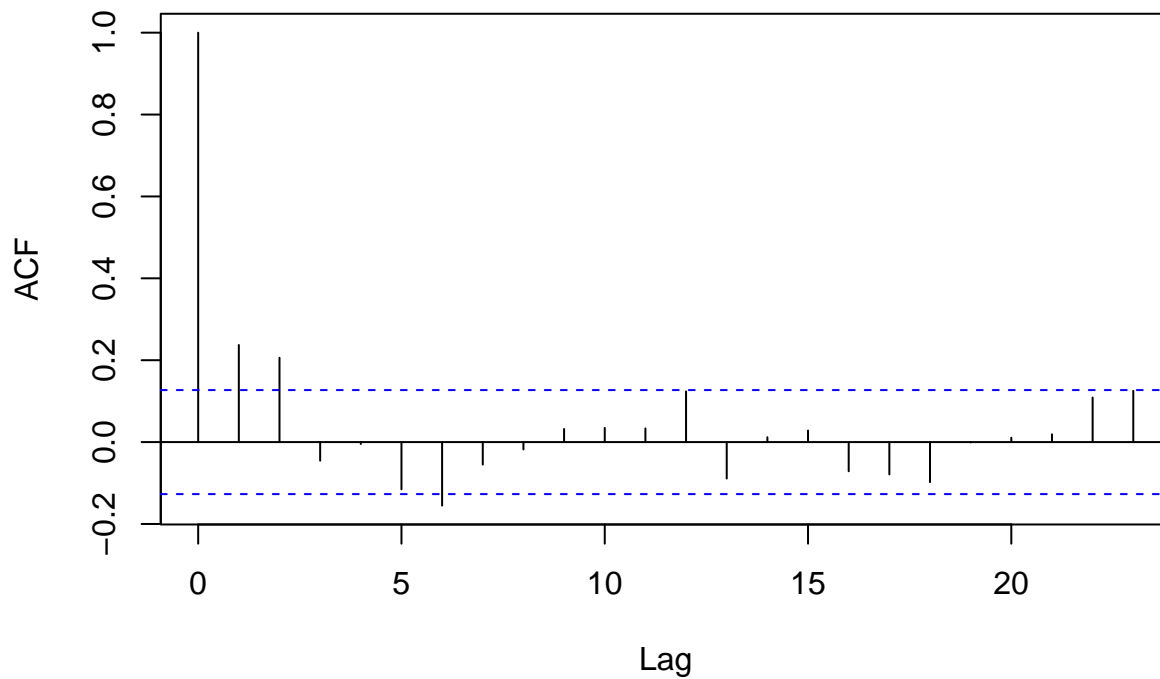
```
esmdata[32]
```

```
## # A tibble: 238 x 1
##   depression_factor
##   <dbl>
## 1          91.3
## 2          10.5
## 3         -16.4
## 4         -20.3
```

```
## 5          29.8
## 6          34.6
## 7         -11.6
## 8          31.7
## 9          14.8
## 10         29.8
## # ... with 228 more rows
```

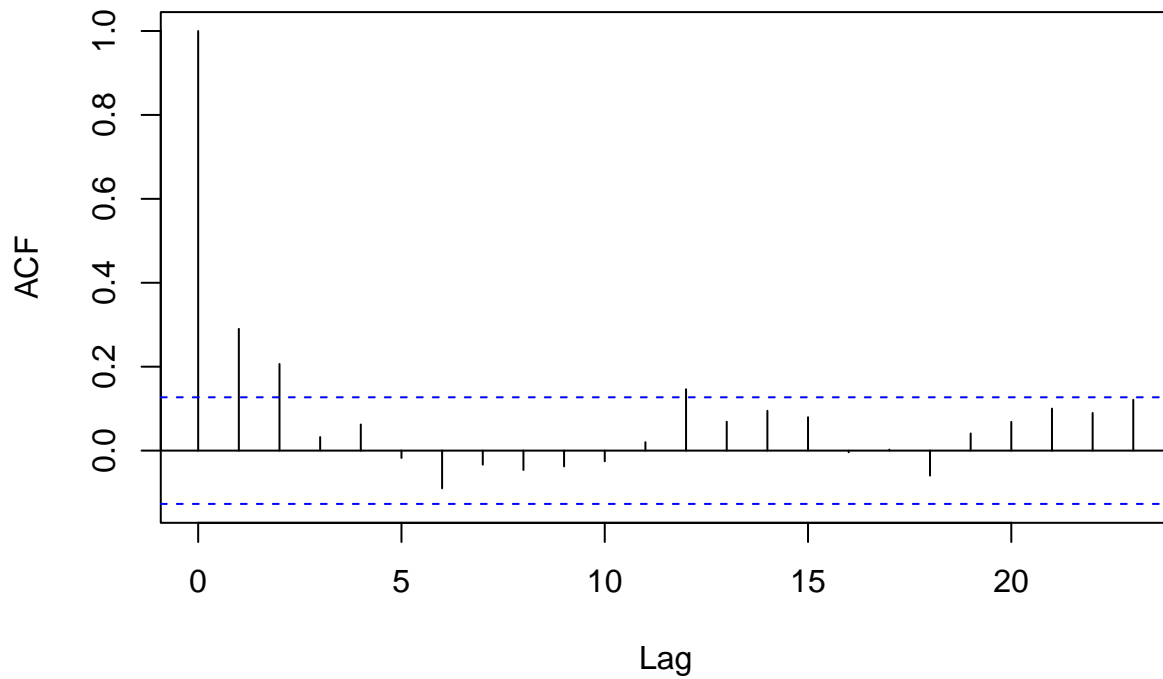
```
acf(esmdata[32], na.action = na.pass)
```

### depression\_factor



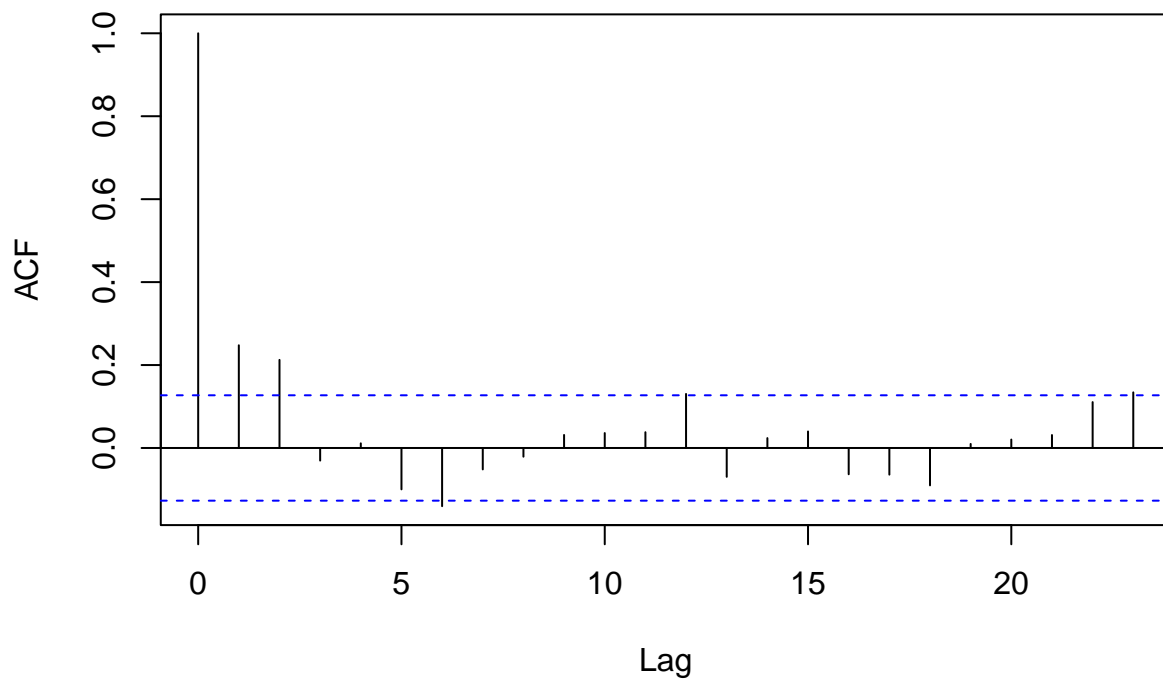
```
# depressive moods
acf(esmdata[31])
```

### depressive\_moods



```
# all moods accumulated  
acf(esmdata[29] + esmdata[30] + esmdata[31])
```

### positive\_moods



► Based on (basic) content knowledge about the variable, and these visualizations, is

there reason to assume the data are non-stationary and/or that there is a seasonal component?

## 2. Forecasting

### 2.1. SARIMA modeling

► Perform the Dickey-Fuller test. What is your conclusion?

```
adf.test(esmdata$depression_factor)
```

```
## Warning in adf.test(esmdata$depression_factor): p-value smaller than printed p-  
## value
```

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: esmdata$depression_factor  
## Dickey-Fuller = -6.4837, Lag order = 6, p-value = 0.01  
## alternative hypothesis: stationary
```

► Fit an (S)ARIMA model to the data; what is the order of the model that was selected?

```
esmdata <- tsibble::fill_gaps(esmdata)  
fit_esmdata <- model(esmdata, ARIMA(depression_factor))
```

```
## Warning in sqrt(diag(best$var.coef)): NaNs produced
```

```
report(fit_esmdata)
```

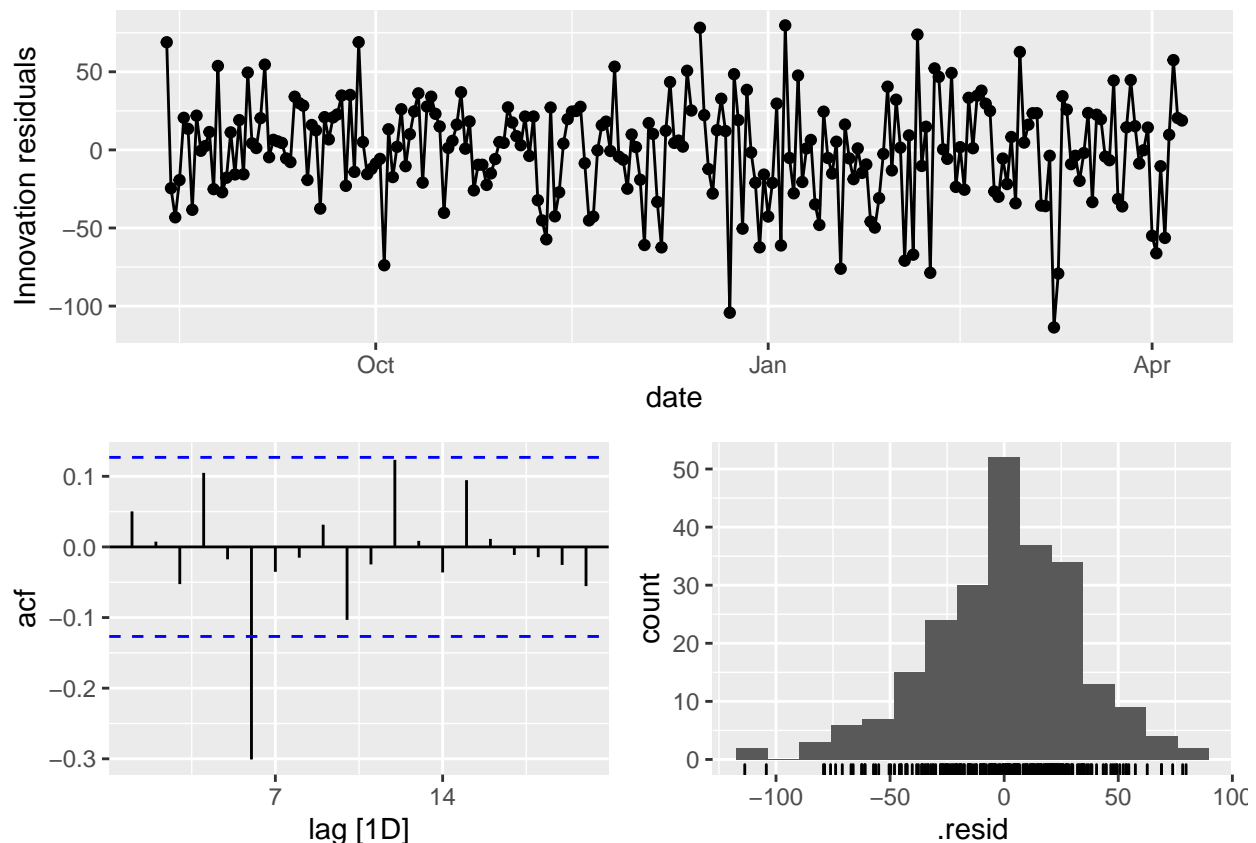
```
## Series: depression_factor  
## Model: ARIMA(1,0,2)(2,0,1)[7] w/ mean  
##  
## Coefficients:  
##          ar1      ma1      ma2      sar1      sar2      sma1  constant  
##      -0.3652  0.5910  0.2971  0.0082  -0.0160  -0.0289   25.9892  
## s.e.   0.2387  0.2322  0.0698      NaN   0.0686      NaN    3.8577  
##  
## sigma^2 estimated as 1082:  log likelihood=-1166.26  
## AIC=2348.52  AICc=2349.15  BIC=2376.33
```

► Check the residuals of the model using the function `gg_tsresiduals()`. What is your conclusion?

```
gg_tsresiduals(fit_esmdata)
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```



```
augment(fit_esmdata)
```

```
## # A tsibble: 239 x 6 [1D]
## # Key:       .model [1]
##   .model          date      depression_factor .fitted .resid .innov
##   <chr>          <date>          <dbl>    <dbl>  <dbl>  <dbl>
## 1 ARIMA(depression_factor) 2012-08-13          91.3    22.4   68.9   68.9
## 2 ARIMA(depression_factor) 2012-08-14          10.5    35.0  -24.5  -24.5
## 3 ARIMA(depression_factor) 2012-08-15         -16.4    26.7  -43.2  -43.2
## 4 ARIMA(depression_factor) 2012-08-16         -20.3   -0.970 -19.3  -19.3
## 5 ARIMA(depression_factor) 2012-08-17          29.8     9.20   20.6   20.6
## 6 ARIMA(depression_factor) 2012-08-18          34.6    21.1   13.5   13.5
## 7 ARIMA(depression_factor) 2012-08-19         -11.6    26.7  -38.4  -38.4
## 8 ARIMA(depression_factor) 2012-08-20          31.7     9.75   21.9   21.9
## 9 ARIMA(depression_factor) 2012-08-21          14.8    15.4   -0.568 -0.568
## 10 ARIMA(depression_factor) 2012-08-22          29.8    27.3    2.42    2.42
## # ... with 229 more rows
```

## 2.2. Dynamic regression

► Include the predictor in an dynamic regression model (i.e., allow for (S)ARIMA residuals); what is the effect of the predictor?

- ▶ What order is the (S)ARIMA model for the residuals?
- ▶ Check the residuals of the model using the function `gg_tsresiduals()`. What is your conclusion?

## 2.3. Forecasts

- ▶ Choose a forecasting horizon, and indicate why this is a reasonable and interesting horizon to consider.
- ▶ Create forecasts based on the model without the predictor and plot these.
- ▶ Create forecasts based on the model with the predictor and plot these.
- ▶ Compare the plots of both forecasts (visually), and discuss how they are similar and/or different.

## 3. Causal Modeling

- ▶ Formulate a causal research question(s) involving the time series variable(s) you have measured.
- ▶ Which method we learned about in class (Granger causal approaches, interrupted time series, synthetic controls) is most appropriate to answer your research question using the data you have available? Why?

### 3.2 Analysis

Depending on the choice you made above, follow the questions outlined in 3.2a, 3.2b or 3.2c. If you chose a Granger causal analysis, it is sufficient to assess Granger causality in one direction only: you may evaluate a reciprocal causal relationship, but then answer each question below for both models.

#### 3.2a Granger Causal analysis

- ▶ Visualize your putative cause variable(s)  $X$  and outcome variables  $Y$ .
- ▶ Train an appropriate ARIMA model on your outcome variable(s)  $Y$ , ignoring the putative cause variable(s) ( $X$ ) but including, if appropriate, any additional covariates. If using the same model as fit in part 2, briefly describe that model again here.
- ▶ Justify what range of lags to consider for the lagged predictor(s). Use the CCF, but you may also justify this based on domain knowledge or substantive theory.
- ▶ Investigate whether adding your lagged “cause” variables ( $X$ ) improve the prediction of your effect variable(s)  $Y$ . Use model selection based on information criteria. Describe your final chosen model

### 3.2b Interrupted Time Series analysis

- ▶ Partition your dataset into pre- and post- intervention time periods and visualize this. Describe what the intervention is and when it takes place
- ▶ Train an appropriate ARIMA model on pre-intervention data of your outcome variable  $Y$ . If using the same model as fit in part 2, briefly describe that model again here.
- ▶ Use this model to create forecasts for the post-intervention time period. Visualize your forecasts (both point predictions and intervals) and the observed post-intervention data in a single plot.
- ▶ Compare your forecasts (both point predictions and intervals) to the observed post-intervention data. Describe if and how these differ from one another.

### 3.2c Synthetic Control Analysis

- ▶ Partition your dataset into pre- and post- intervention time periods. Describe what the intervention is and when it takes place. Describe your control series. Visualize your original time-series, control series, and the intervention period.
- ▶ Train an appropriate model on pre-intervention data of your outcome variable and the control series. Describe the model. Note: if using *CausalImpact*, describe the fitted model before visualizing the forecasts.
- ▶ Use this model to create forecasts for the post-intervention time period. Visualize your forecasts (both point predictions and intervals) and the observed post-intervention data in a single plot.
- ▶ Compare your forecasts (both point predictions and intervals) to the observed post-intervention data. Describe if and how these differ from one another.

### 3.3 Conclusion and critical reflection

- ▶ Based on the result of your analysis, how would you answer your causal research question?
  - ▶ Making causal conclusions on the basis of your analysis is reliant on a number of assumptions. Pick a single assumption that is necessary in the approach you chose. Discuss the plausability and possible threats to the validity of this assumption in your specific setting (< 75 words)
-