# Flowering exploration of the 1001G fiel experiment

*Moises Exposito-Alonso (moisesexpositoalonso@gmail.com)*

*2017-06-01*

### Packages set up

```
library(knitr)
library(dplyr,tidyr)
library(ggplot2);library(cowplot)
library(devtools)
library(RColorBrewer)
library(stargazer)
library(moiR)
#source("~/ebio/abt6_projects9/ath_1001G_image_pheno/experiment_218_droughtgwa/droughtfunctions.R")
#source("droughtfunctions.R")
load_all(".") # field
```

### load dataset

```
data(field)
dim(field)
```

```
## [1] 24747    22
```

```
names(field)
```

```
##  [1] "qp"         "pos"        "site"       "id"         "rep"
##  [6] "trayid"     "indpop"     "water"      "name"       "country"
## [11] "latitude"   "longitude"  "kgroup"     "FT.q"       "FT.date"
## [16] "FT.dif"     "pathimage"  "folder"     "image"      "count"
## [21] "num"        "Harv.q"
```

```
stargazer(field,type='text')
```

```
##
## ======================================================
## Statistic   N      Mean    St. Dev.    Min      Max
## ------------------------------------------------------
## id        24,747 8,282.194 1,997.319    1     10,020
## rep       24,747   3.578     1.843      1        7
## latitude  24,651  47.524     7.367    15.111  63.083
## longitude 24,651  10.552    24.673   -119.350 136.310
## kgroup    24,315   5.849     3.112      1       11
## FT.q      24,747   0.680     0.486     -9        1
## FT.dif    16,858  139.555   25.180     64       178
## Harv.q    24,747   0.000     0.000      0        0
## ------------------------------------------------------
```

**How frequent flowering was measured**

```
data("flowering_madrid")
uniqueflowering<-sort(unique(flowering_madrid$FT.date)  ) - sowingday('madrid')
difflowering<-diff(uniqueflowering)
summary(fn(difflowering))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   1.839   2.000  13.000
```

```
data("flowering_madrid")
uniqueflowering<-sort(unique(flowering_tuebingen$FT.date)  ) - sowingday('tuebingen')
difflowering<-diff(uniqueflowering)
summary(fn(difflowering))
```

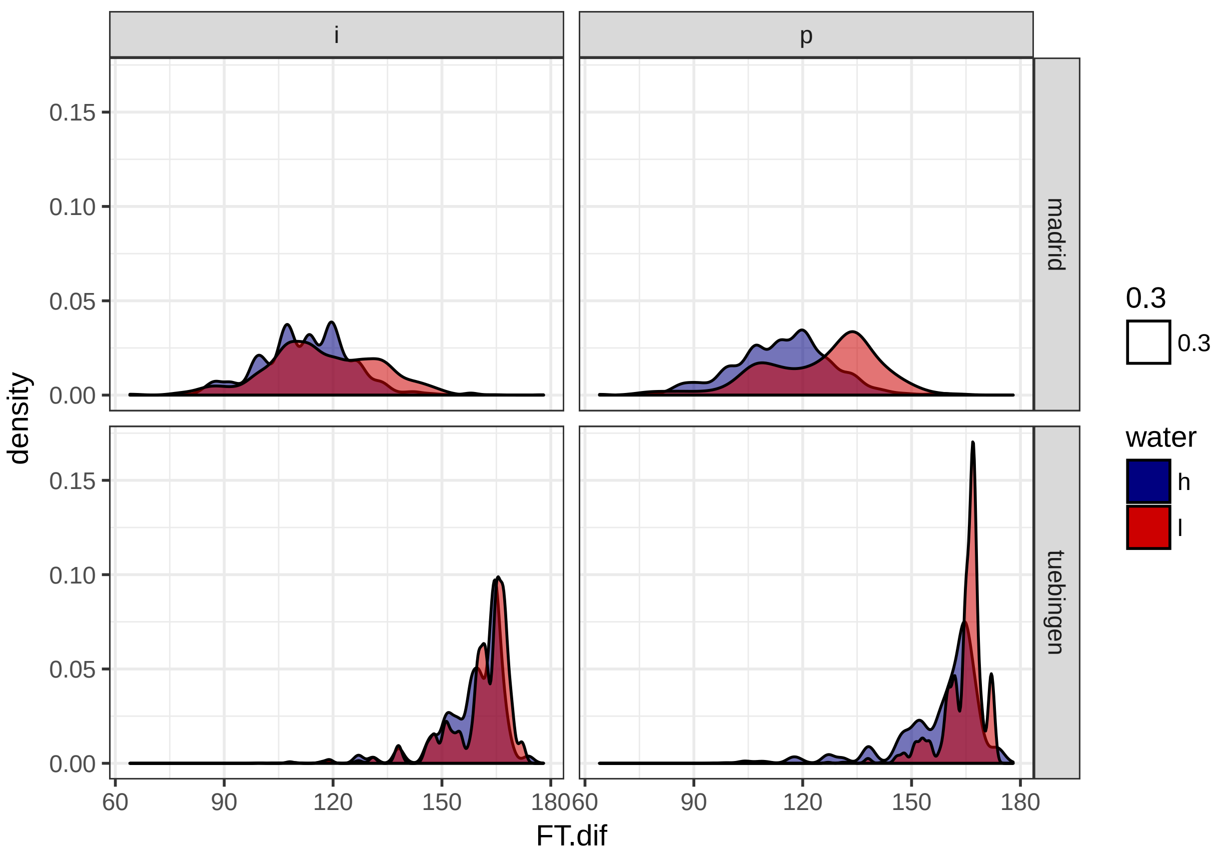
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   2.000   2.581   3.000   8.000
```

Flowering was recorded on average ever 1 day, while in Tubingen every 2 days.

## Plot flowering histograms

```
p1=ggplot(data = field)+
  geom_density(aes(x=FT.dif,group=water,fill=water,alpha=0.3)) + scale_fill_manual(values = c("h"="navy
p1
```

```
## Warning: Removed 7889 rows containing non-finite values (stat_density).
```



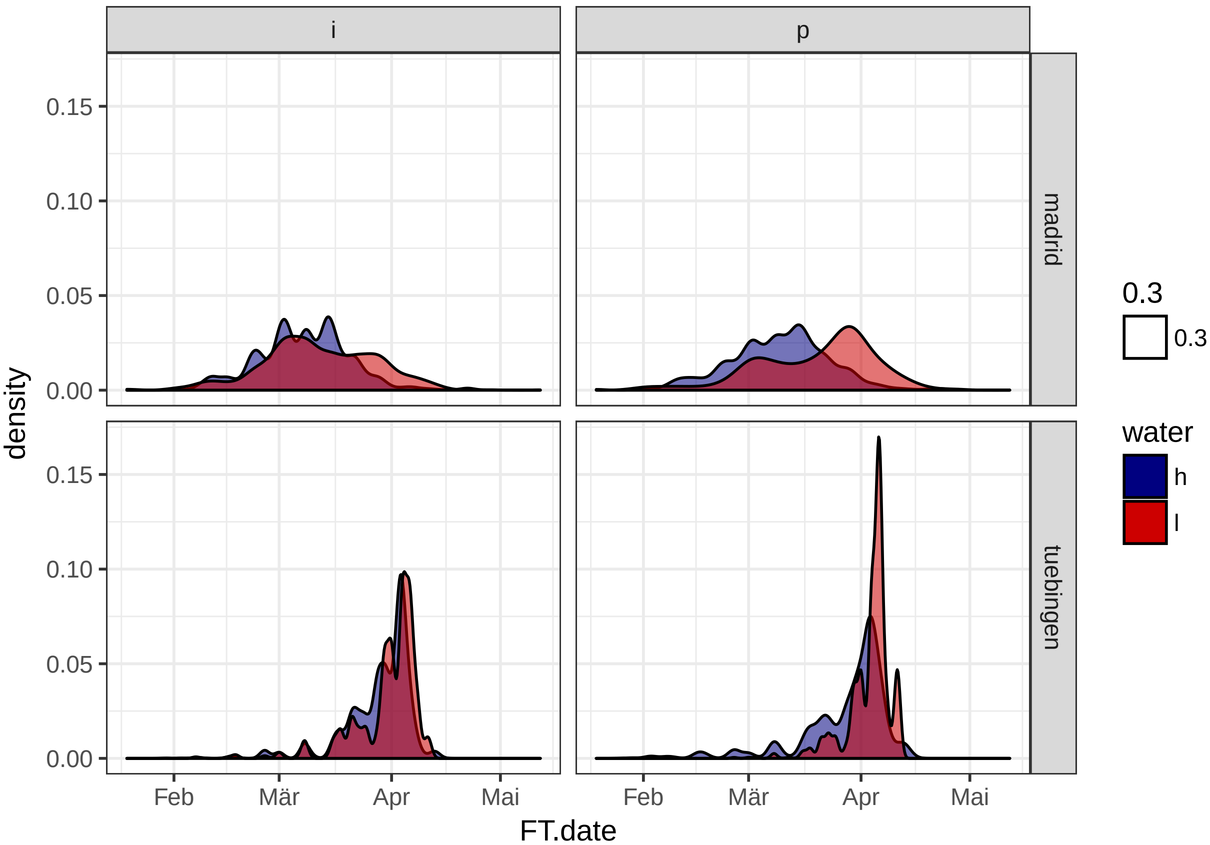
Histograms of the flowering time from the day of harvesting until the recorded day per replicate.

The first observation is that in Tuebingen there is more variance in flowering time. In Madrid flowerin time is in a narrow peak. Maybe it indicates that later than early april everything would die.

The second observation is that under low water conditions it seems that the mean flowering time is later. This can be seen as opposite to what the field expects, but it might not be necessarily true. Normally, when a new stress comes to a plant, they try to accelerate the phenology. However, if the stress is persistent since the start of the experiment, they might not be able to reach the necessary developmental stage to be able to flower.

```
p1a=ggplot(data = field)+
  geom_density(aes(x=FT.date,group=water,fill=water,alpha=0.3)) + scale_fill_manual(values = c("h"="navy
p1a
```

```
## Warning: Removed 7889 rows containing non-finite values (stat_density).
```

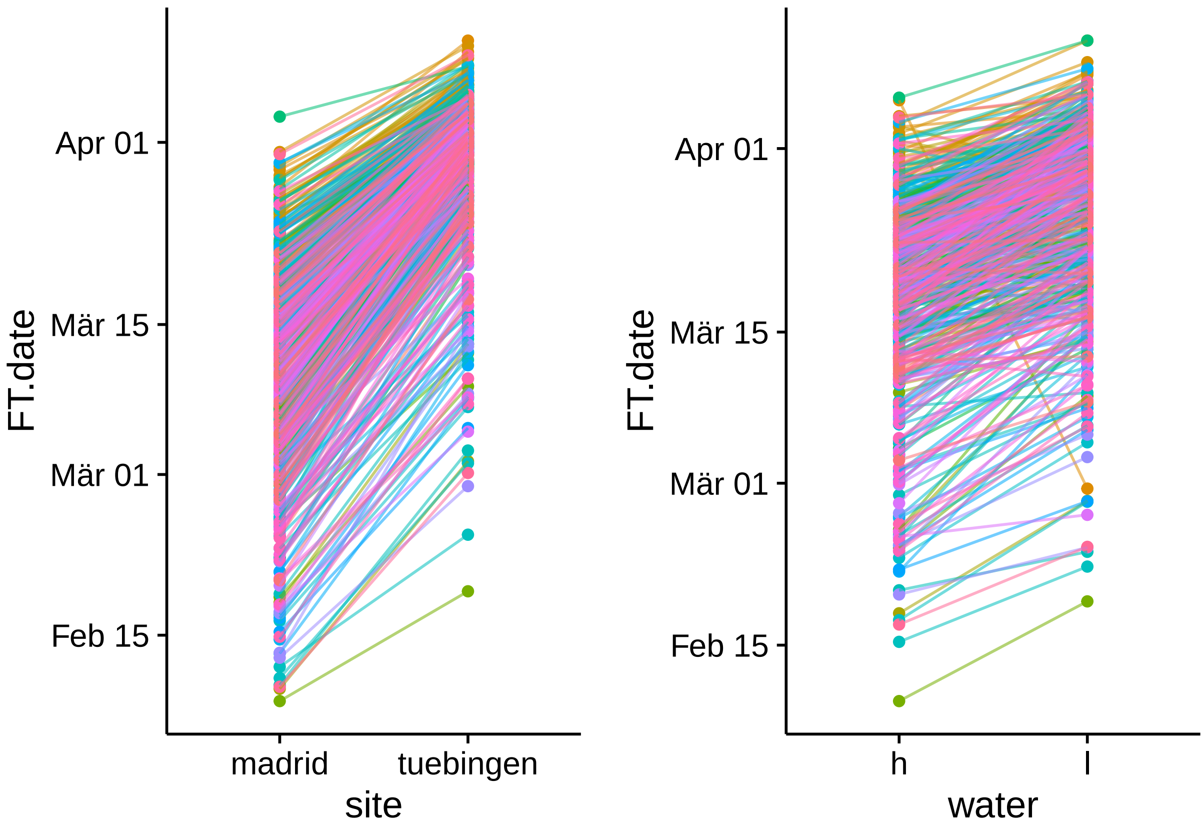



Histograms of the exact date of flowering (month and day). A similar picture can be drawn from these plots as above. However, in this case we can actually see the dates of flowering. Apparently, in Spain the dates are towards April compared to Tuebingen that spam from februrry to april. Perhaps due to the fact that we started the experiment in Tuebingen 3 weeks earlier than in Madrid.

### Two sites or two treatments lines to visualize interaction

```
p3a=field %>% group_by(id,site) %>% summarise(.,FT.date=mean(FT.date,na.rm=T))  %>%
  ggplot(., aes(y=FT.date,x=site, group=id,color=factor(id))) + geom_point() +
  geom_line(aes(alpha=0.5)) + theme(legend.position="none")
```

```r
p4a=field %>% group_by(id,water) %>% summarise(.,FT.date=mean(FT.date,na.rm=T))  %>%
  ggplot(., aes(y=FT.date,x=water, group=id,color=factor(id)))  + geom_point() +
  geom_line(aes(alpha=0.5)) + theme(legend.position="none")

plot_grid(p3a,p4a, rel_widths = c(2,2))
```




This plot, which is based in means per genotype and per site, shows again that in Tuebingen plants flowered in an earlier date generally. The plot in the right indicates that under lower water content, plants flowered later. However, in this case it seems that the interaction is a bit larger, that is, there is more lines that are crossing.

```r
subsetfield=field %>% filter(., FT.date < as.Date('2016-01-20' ,format= "%Y-%m-%d")  ) %>% select(., id
dim(subsetfield)
```

```
## [1] 11  8
```

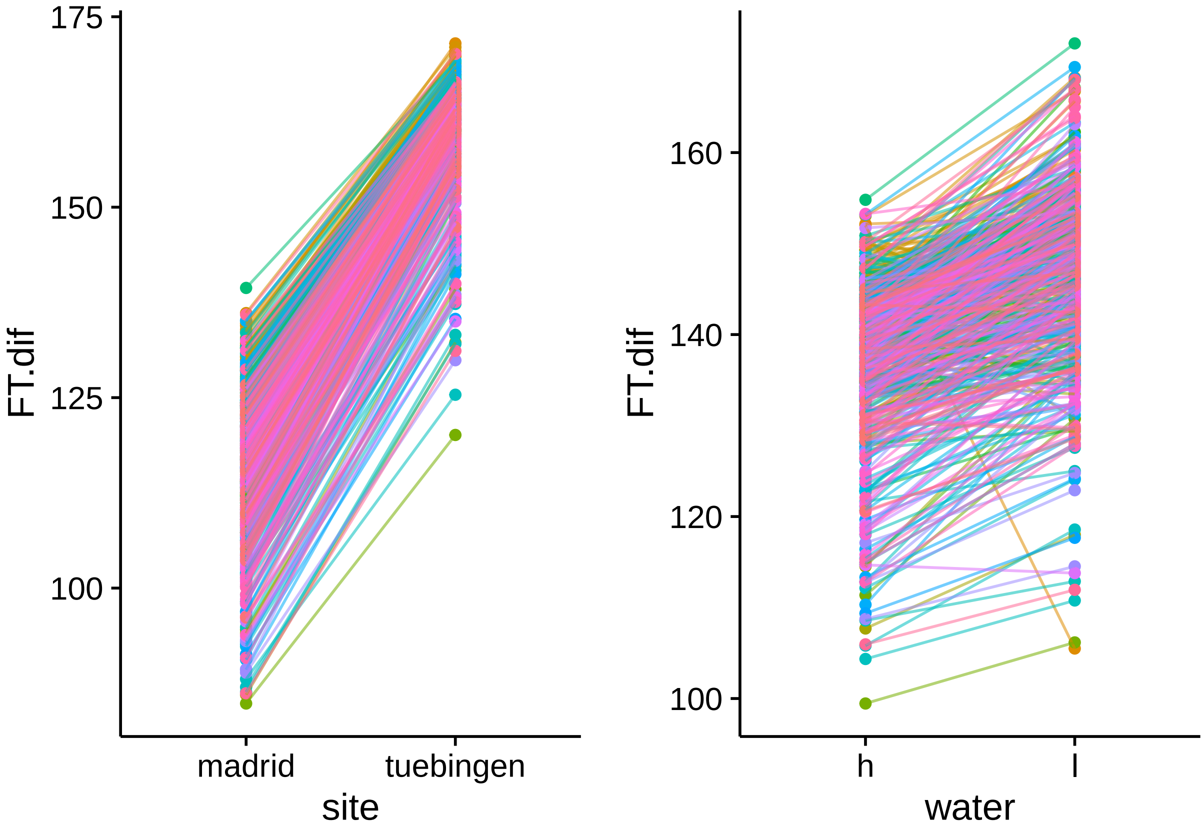
```r
print(subsetfield)
```

```
##      id   site water indpop    FT.date kgroup    name country
## 1  6099 madrid     l      i 2016-01-19      3   T1090     SWE
## 2  9855 madrid     l      i 2016-01-19      4   Lam-0     ESP
## 3  7169 madrid     l      i 2016-01-19      9    Hh-0     GER
## 4  9941 madrid     l      i 2016-01-19     NA   Fei-0     POR
## 5  9870 madrid     h      p 2016-01-19     10   Moz-0     ESP
## 6  7343 madrid     h      p 2016-01-19      3    Sp-0     GER
## 7  9517 madrid     h      p 2016-01-19      4 IP-All-0     ESP
## 8  9527 madrid     h      p 2016-01-19      4 IP-Cad-0     ESP
## 9  9512 madrid     h      p 2016-01-19      8 IP-Vid-1     POR
## 10 8243 madrid     l      i 2016-01-19      9   PHW-2     ITA
```

```
## 11 6148 madrid     l       i 2016-01-19      6      T960      SWE
```

Interesting that the ones with the earliset flowering time, in January in Tuebingen, were actually in low water and from the two extremes of the distribution, Spain and Sweden

```
p3=field %>% group_by(id,site) %>% summarise(.,FT.dif=mean(FT.dif,na.rm=T))  %>%
  ggplot(., aes(y=FT.dif,x=site, group=id,color=factor(id)))  + geom_point() +
  geom_line(aes(alpha=0.5)) + theme(legend.position="none")

p4=field %>% group_by(id,water) %>% summarise(.,FT.dif=mean(FT.dif,na.rm=T))  %>%
  ggplot(., aes(y=FT.dif,x=water, group=id,color=factor(id))) + geom_point() +
  geom_line(aes(alpha=0.5)) + theme(legend.position="none")

plot_grid(p3,p4, rel_widths = c(2,2))
```

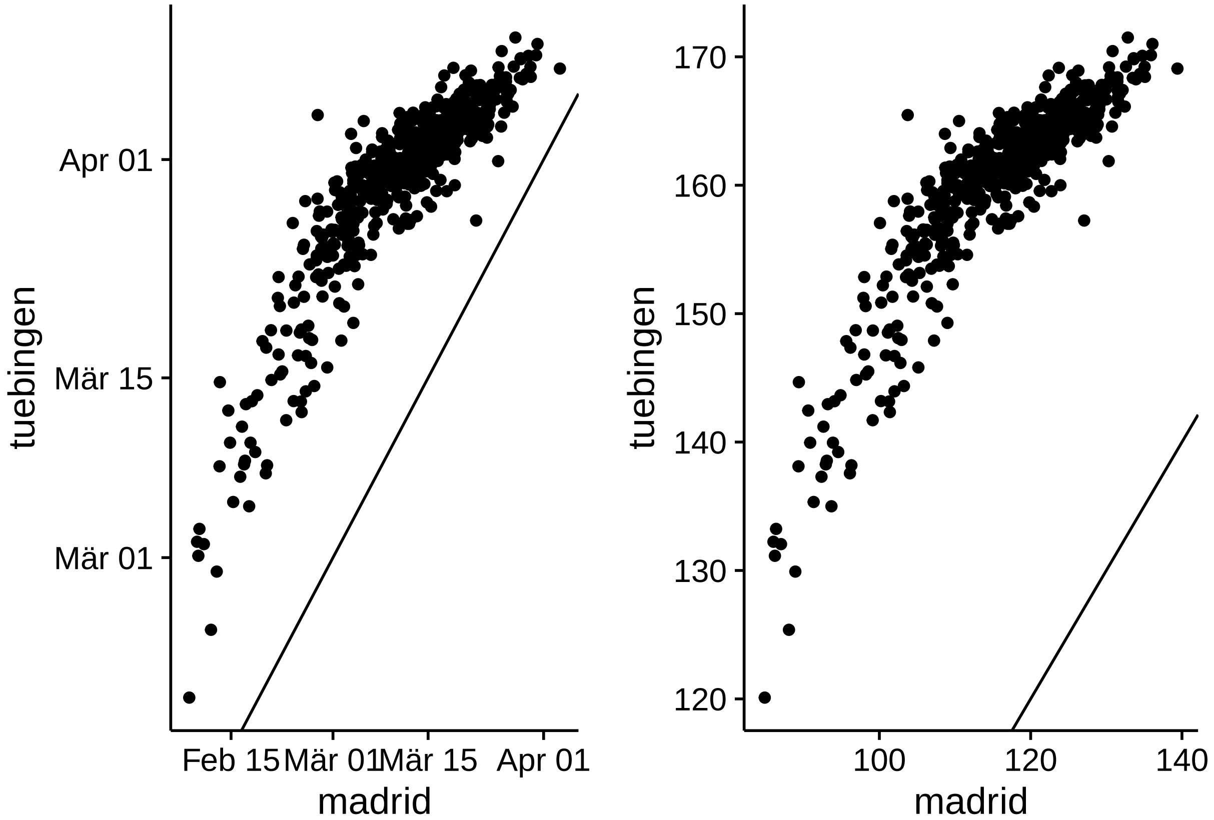



With the flowering time relative to the date of germination (instead of day and month of flwoering ), we see that Tuebingen was later (opposite trend, but we know that it is because we germinated Madrid later so they had to be faster to live) and low water was later (same trend). Nice that again we see interaction, some lines go in one direction and other in the opposite.

### Scatter plots genotype means between treatments

```
p2a= field %>% group_by(site,id) %>% summarise(.,flowering=mean(FT.date,na.rm=T)) %>% tidyr::spread(., k
  ggplot(data = .)+ geom_point(aes(x=madrid,y=tuebingen)) + geom_abline(aes(intercept=0,slope=1))

p2= field %>% group_by(site,id) %>% summarise(.,flowering=mean(FT.dif,na.rm=T)) %>% tidyr::spread(., key
  ggplot(data = .)+ geom_point(aes(x=madrid,y=tuebingen)) + geom_abline(aes(intercept=0,slope=1))
```

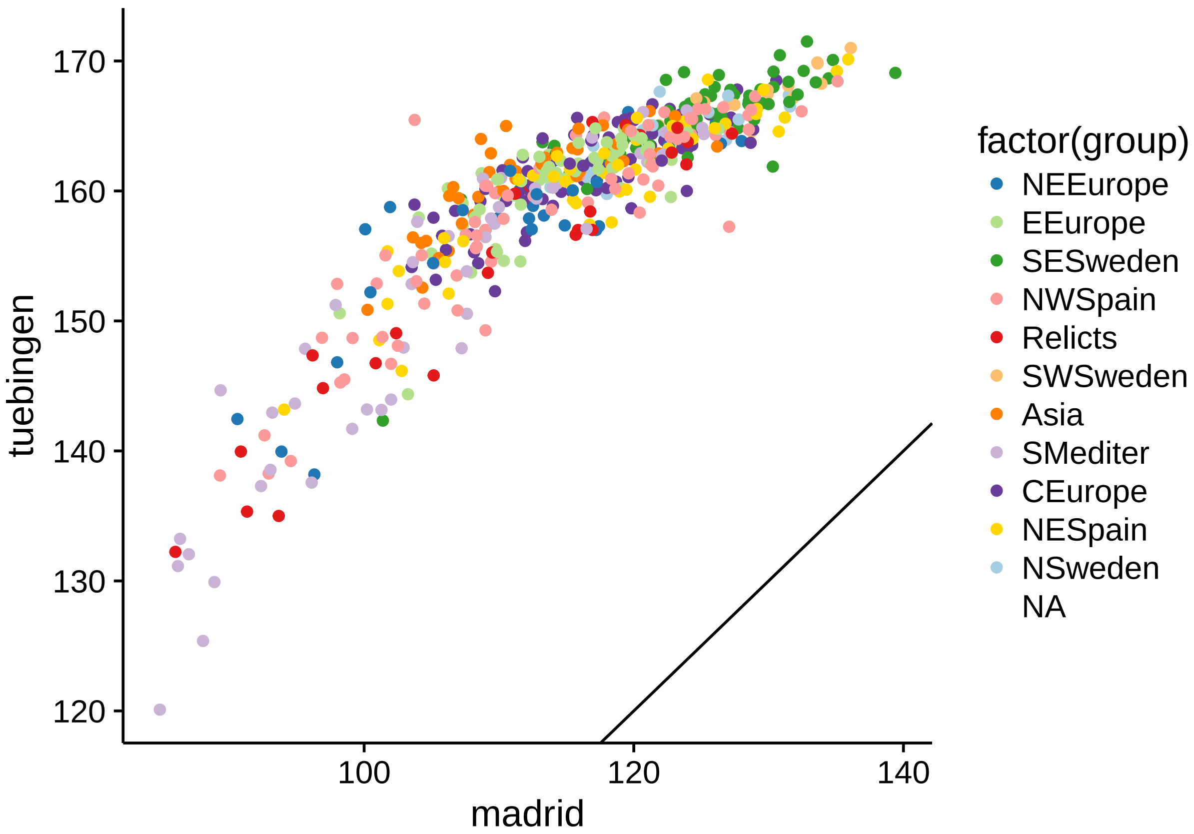
```
plot_grid(p2a,p2,rel_widths = c(2,2))
```




As final comparison, two scatter plots are shown with the mean date (left) and time (right) per genotype are correlated in both locations. This was expected and indicates that the rank of early to late flowering genotypes is more or less maitained, but that the extremes are more dispersed in Tuebingen (we mentioned the variance before), and this is visualized with the ends in U-shape.

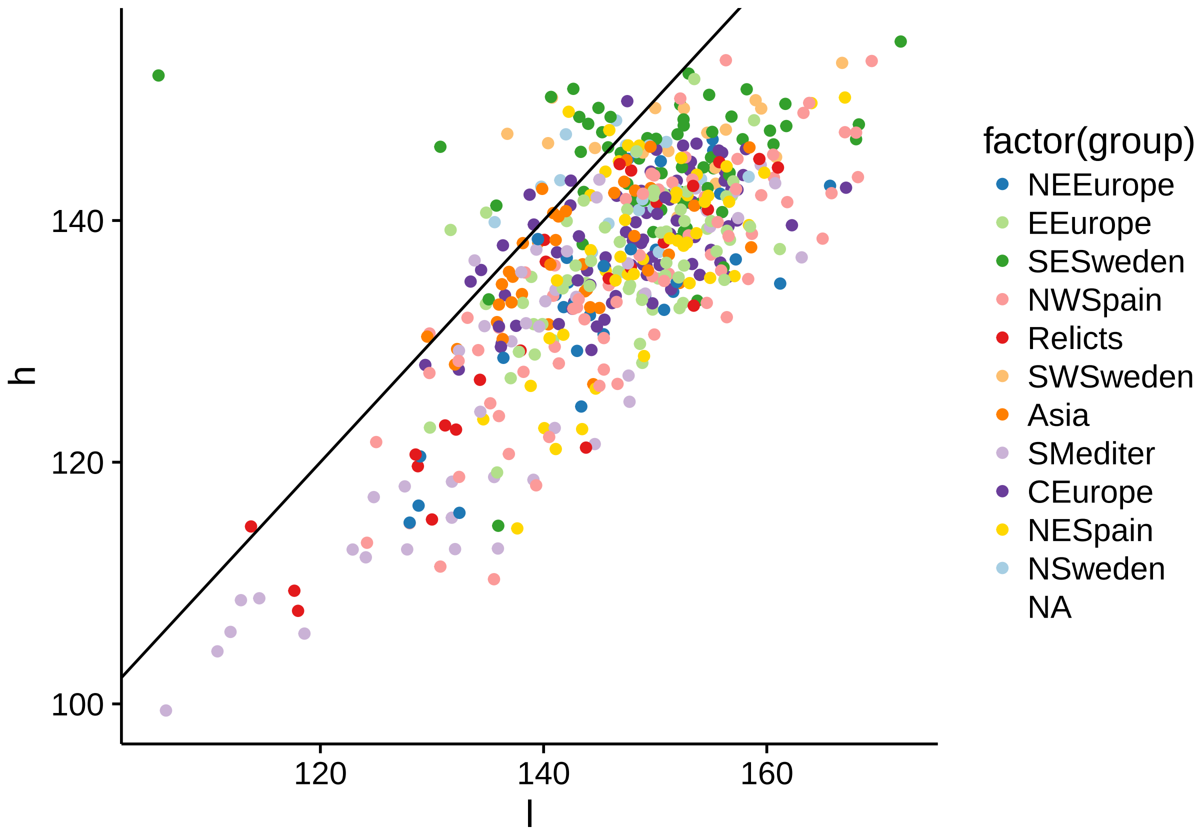
```
field$kgroup<-factor(field$kgroup,labels = knames())

p3= field %>% group_by(site,id) %>% summarise(.,flowering=mean(FT.dif,na.rm=T), group= unique(kgroup)) %
  ggplot(data = .)+ geom_point(aes(x=madrid,y=tuebingen, color=factor(group),group=group)) + geom_ablin
p3<-p3  +scale_colour_manual(values = colors11())

p3
```

```
## Warning: Removed 9 rows containing missing values (geom_point).
```




```
# p<-createmapbase()+coord_equal(xlim =  c(-26,+95 ), ylim=c(10,70))
# p+ geom_point(data=field,aes(y=latitude,x=longitude),color="black",size=2)  + geom_point(data=field,a
# p
p4= field %>% group_by(water,id) %>% summarise(.,flowering=mean(FT.dif,na.rm=T), group= unique(kgroup))
  ggplot(data = .)+ geom_point(aes(x=l,y=h, color=factor(group),group=group)) + geom_abline(aes(intercep
p4<-p4  +scale_colour_manual(values = colors11())
p4
```

```
## Warning: Removed 9 rows containing missing values (geom_point).
```




```
## now plot of the differences
# difs=field %>% group_by(site,water,id) %>% summarise(.,flowering=mean(FT.dif,na.rm=T), group= unique(
#   spread(WaterSite,value=flowering) %>%
  # mutate(high_low=h_madrid-l_madrid,tue_mad=)


# difs %>%
#   ggplot(data = .)+ geom_point(aes(x=l,y=h, color=factor(group),group=group)) + geom_abline(aes(inter
# p5<-p5  +scale_colour_manual(values = colors11())
```

From this last plot, we see where the accessions that flower at different times come from. Nice to see that the mediterraneans (lila) and relicts (red) thend to be earliset, whereas ome from Swedish (dark green) and from the Pirinees (yellow).

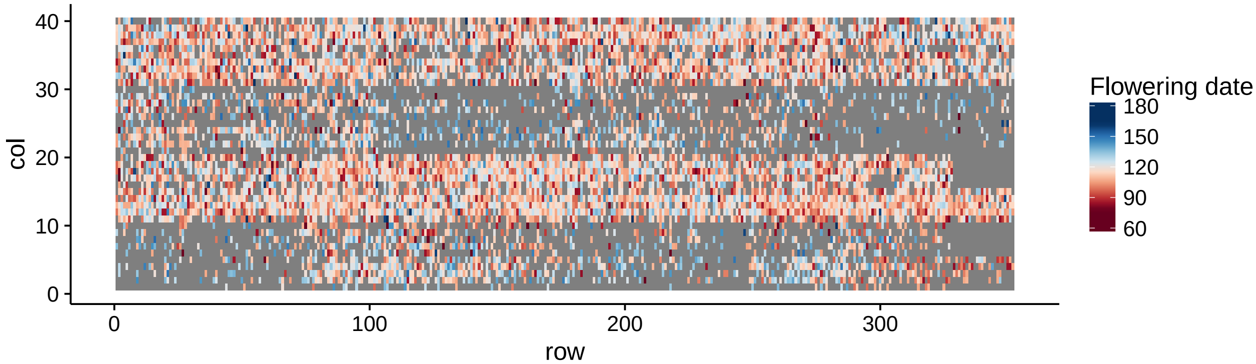
## Direct visualization of flowering time in the spatial positions of the experiment

```
flo=read_flowering('madrid','../data-raw')
```

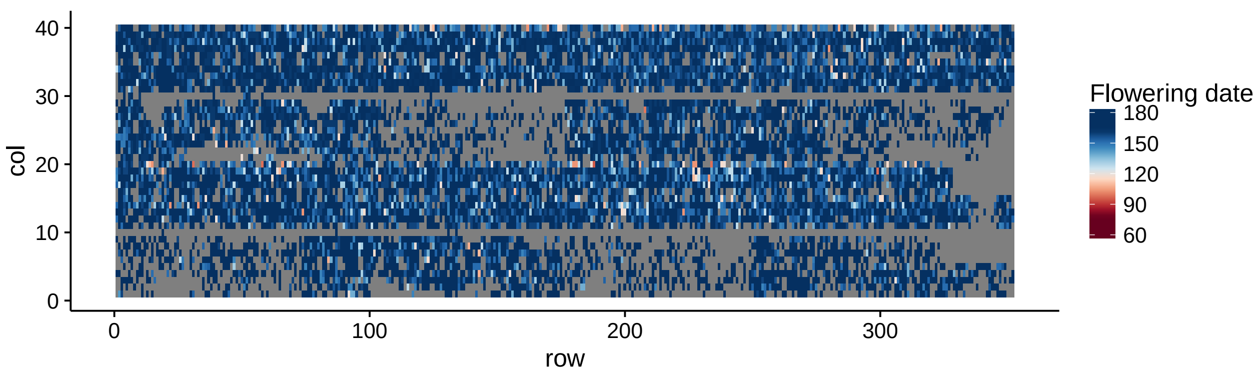
```
## [1] "../data-raw/Flowering_pheno_Madrid.xlsx-combined.tsv"
```

```
flo=as.matrix(flo)
flo=apply(flo,2,function(x)asmydate(x,'madrid'))
flo[traycoordinates_tunnel() == "x"] <-NA # Remove those corners

f=data.frame(row=fn(row(flo)),col=fn(col(flo)),flowering=fn(flo))
```

```
f1=ggplot(f,aes(x=row,y=col , fill=flowering ))+geom_tile() +
  scale_fill_gradientn("Flowering date",
                      colours= make.pallete.contrast( vecol = brewer.pal(10,name = "RdBu"),contrast=2,
                       limits=c(60, 180)
)

f1
```




```
flo=read_flowering('tuebingen','../data-raw')
flo=as.matrix(flo)
flo=apply(flo,2,function(x)asmydate(x,'tuebingen'))
flo[traycoordinates_tunnel() == "x"] <-NA # Remove those corners

f=data.frame(row=fn(row(flo)),col=fn(col(flo)),flowering=fn(flo))

f1=ggplot(f,aes(x=row,y=col , fill=flowering ))+geom_tile() +
  scale_fill_gradientn("Flowering date",
                      colours= make.pallete.contrast( vecol = brewer.pal(10,name = "RdBu"),contrast=2,
                      limits=c(60, 180)
                      )

f1
```




```
# summary(c(flo))
```

We see that Tuebingen again has later flowering times than Madrid. This also helps to visualize that the blocks of lower watering had

**load data sets**

```
data(field)
dim(field)
```

```
## [1] 24747    22
```

```
names(field)
```

```
##  [1] "qp"        "pos"       "site"      "id"        "rep"
##  [6] "trayid"    "indpop"    "water"     "name"      "country"
## [11] "latitude"  "longitude" "kgroup"    "FT.q"      "FT.date"
## [16] "FT.dif"    "pathimage" "folder"    "image"     "count"
## [21] "num"       "Harv.q"
```

```
stargazer(field,type='text')
```

```
##
## =====================================================
## Statistic   N      Mean    St. Dev.    Min      Max
## -----------------------------------------------------
## id        24,747 8,282.194 1,997.319    1     10,020
## rep       24,747   3.578     1.843      1        7
## latitude  24,651  47.524     7.367    15.111  63.083
## longitude 24,651  10.552    24.673   -119.350 136.310
## kgroup    24,315   5.849     3.112      1       11
## FT.q      24,747   0.680     0.486      -9       1
## FT.dif    16,858  139.555   25.180      64      178
## Harv.q    24,747   0.000     0.000      0        0
## -----------------------------------------------------
```

## Fixed Model

```
library(stargazer)
library(MCMCglmm)
```

```
## Loading required package: Matrix

## Loading required package: coda

## Loading required package: ape
```

```
names(field)
```

```
##  [1] "qp"        "pos"       "site"      "id"        "rep"
##  [6] "trayid"    "indpop"    "water"     "name"      "country"
## [11] "latitude"  "longitude" "kgroup"    "FT.q"      "FT.date"
## [16] "FT.dif"    "pathimage" "folder"    "image"     "count"
## [21] "num"       "Harv.q"
```

```
lmod<-lm(data=field,
  FT.dif ~  latitude
   )
lmod2<-lm(data=field,
  FT.dif ~  latitude + longitude^2
   )
```

```r
stargazer(lmod,lmod2,type='text')
```

```
##
## ========================================================================
##                                     Dependent variable:
##                     ----------------------------------------------------
##                                           FT.dif
##                                (1)                        (2)
## ------------------------------------------------------------------------
## latitude                    0.511***                   0.521***
##                             (0.026)                    (0.027)
##
## longitude                                               -0.013
##                                                        (0.008)
##
## Constant                   115.361***                 115.008***
##                             (1.248)                    (1.267)
##
## ------------------------------------------------------------------------
## Observations                 16,794                     16,794
## R2                           0.022                      0.023
## Adjusted R2                  0.022                      0.022
## Residual Std. Error    24.903 (df = 16792)        24.901 (df = 16791)
## F Statistic         384.261*** (df = 1; 16792) 193.440*** (df = 2; 16791)
## ========================================================================
## Note:                                        *p<0.1; **p<0.05; ***p<0.01
```

### Mixed Model

```r
library(lme4)
library(lmerTest)
```

```
##
## Attaching package: 'lmerTest'

## The following object is masked from 'package:lme4':
##
##     lmer

## The following object is masked from 'package:stats':
##
##     step
```

```r
library(stargazer)
library(xtable)
data(cake)

# Get the table first.
summary(M1 <- lme4::lmer(angle ~ temp + (1 | replicate) + (1|recipe:replicate), cake, REML= FALSE))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: angle ~ temp + (1 | replicate) + (1 | recipe:replicate)
##    Data: cake
```

```
##
##      AIC      BIC   logLik deviance df.resid
##   1668.8   1686.7   -829.4   1658.8      265
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.77207 -0.56112 -0.03137  0.57250  2.64880
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  recipe:replicate (Intercept)  3.973   1.993
##  replicate        (Intercept) 35.324   5.943
##  Residual                     20.615   4.540
## Number of obs: 270, groups:  recipe:replicate, 45; replicate, 15
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  0.51587    3.60429   0.143
## temp         0.15803    0.01618   9.767
##
## Correlation of Fixed Effects:
##      (Intr)
## temp -0.898
```

```
summary(M2 <- lme4::lmer(angle ~ factor(temperature) + (1 | replicate) + (1|recipe:replicate), cake, REI
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## angle ~ factor(temperature) + (1 | replicate) + (1 | recipe:replicate)
##    Data: cake
##
##      AIC      BIC   logLik deviance df.resid
##   1670.2   1702.6   -826.1   1652.2      261
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.85441 -0.61426 -0.07482  0.55722  2.85207
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  recipe:replicate (Intercept)  4.072   2.018
##  replicate        (Intercept) 35.324   5.943
##  Residual                     20.022   4.475
## Number of obs: 270, groups:  recipe:replicate, 45; replicate, 15
##
## Fixed effects:
##                       Estimate Std. Error t value
## (Intercept)            32.1222     1.5873  20.237
## factor(temperature).L   6.6109     0.6670   9.911
## factor(temperature).Q  -0.3855     0.6670  -0.578
## factor(temperature).C  -0.5483     0.6670  -0.822
## factor(temperature)^4  -1.2977     0.6670  -1.945
## factor(temperature)^5  -0.9141     0.6670  -1.370
##
## Correlation of Fixed Effects:
```

```
##             (Intr) fc().L fc().Q fc().C fc()^4
## fctr(tmp).L 0.000
## fctr(tmp).Q 0.000  0.000
## fctr(tmp).C 0.000  0.000  0.000
## fctr(tmp)^4 0.000  0.000  0.000  0.000
## fctr(tmp)^5 0.000  0.000  0.000  0.000  0.000
```

```
stargazer(M1, M2, style="ajps", title="An Illustrative Model Using Cake Data", dep.var.labels.include =
      covariate.labels=c( "Temperature (Continuous)", "Temperature (Factor $<$ 185)", "Temperature (Fac
)
```

```
##
## % Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvar
## % Date and time: Do, Jun 01, 2017 - 21:17:43
## \begin{table}[!htbp] \centering
##   \caption{An Illustrative Model Using Cake Data}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lcc}
## \\[-1.8ex]\hline \\[-1.8ex]
## \\[-1.8ex] & \textbf{Model 1} & \textbf{Model 2}\\
## \hline \\[-1.8ex]
##  Temperature (Continuous) & 0.158$^{***}$ &  \\
##   & (0.016) &  \\
##   Temperature (Factor $<$ 185) &  & 6.611$^{***}$ \\
##   &  & (0.667) \\
##   Temperature (Factor $<$ 195) &  & $-$0.386 \\
##   &  & (0.667) \\
##   Temperature (Factor $<$ 205) &  & $-$0.548 \\
##   &  & (0.667) \\
##   Temperature (Factor $<$ 215) &  & $-$1.298$^{*}$ \\
##   &  & (0.667) \\
##   Temperature (Factor $<$ 225) &  & $-$0.914 \\
##   &  & (0.667) \\
##   Constant & 0.516 & 32.122$^{***}$ \\
##   & (3.604) & (1.587) \\
##  N & 270 & 270 \\
## Log Likelihood & $-$829.378 & $-$826.090 \\
## AIC & 1668.755 & 1670.180 \\
## BIC & 1686.747 & 1702.565 \\
## \hline \\[-1.8ex]
## \multicolumn{3}{l}{$^{***}$p $<$ .01; $^{**}$p $<$ .05; $^{*}$p $<$ .1} \\
## \end{tabular}
## \end{table}
```

```
# now for lmerTest
summary(M1a <- lmer(angle ~ temp + (1 | replicate) + (1|recipe:replicate), cake, REML= FALSE))
```

```
## Linear mixed model fit by maximum likelihood t-tests use Satterthwaite
##   approximations to degrees of freedom [lmerMod]
## Formula: angle ~ temp + (1 | replicate) + (1 | recipe:replicate)
##    Data: cake
##
##      AIC      BIC   logLik deviance df.resid
##   1668.8   1686.7   -829.4   1658.8      265
##
## Scaled residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -2.77207 -0.56112 -0.03137  0.57250  2.64880
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  recipe:replicate (Intercept)  3.973   1.993
##  replicate        (Intercept) 35.324   5.943
##  Residual                     20.615   4.540
## Number of obs: 270, groups:  recipe:replicate, 45; replicate, 15
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)   0.51587    3.60429 185.34000   0.143    0.886
## temp          0.15803    0.01618 225.00000   9.767   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## temp -0.898
```

```
summary(M2a <- lmer(angle ~ factor(temperature) + (1 | replicate) + (1|recipe:replicate), cake, REML= FA
```

```
## Linear mixed model fit by maximum likelihood t-tests use Satterthwaite
##   approximations to degrees of freedom [lmerMod]
## Formula:
## angle ~ factor(temperature) + (1 | replicate) + (1 | recipe:replicate)
##    Data: cake
##
##      AIC      BIC   logLik deviance df.resid
##   1670.2   1702.6   -826.1   1652.2      261
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.85441 -0.61426 -0.07482  0.55722  2.85207
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  recipe:replicate (Intercept)  4.072   2.018
##  replicate        (Intercept) 35.324   5.943
##  Residual                     20.022   4.475
## Number of obs: 270, groups:  recipe:replicate, 45; replicate, 15
##
## Fixed effects:
##                       Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)            32.1222     1.5873  15.0000  20.237 2.66e-12 ***
## factor(temperature).L   6.6109     0.6670 225.0000   9.911  < 2e-16 ***
## factor(temperature).Q  -0.3855     0.6670 225.0000  -0.578    0.564
## factor(temperature).C  -0.5483     0.6670 225.0000  -0.822    0.412
## factor(temperature)^4  -1.2977     0.6670 225.0000  -1.945    0.053 .
## factor(temperature)^5  -0.9141     0.6670 225.0000  -1.370    0.172
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
```

```
##             (Intr) fc().L fc().Q fc().C fc()^4
## fctr(tmp).L 0.000
## fctr(tmp).Q 0.000  0.000
## fctr(tmp).C 0.000  0.000  0.000
## fctr(tmp)^4 0.000  0.000  0.000  0.000
## fctr(tmp)^5 0.000  0.000  0.000  0.000  0.000
```

```r
anovadf <- data.frame(anova(M1a,M2a))
xtable(anovadf)
```

```
## % latex table generated in R 3.3.2 by xtable 1.8-2 package
## % Thu Jun  1 21:17:44 2017
## \begin{table}[ht]
## \centering
## \begin{tabular}{rrrrrrrrr}
##   \hline
##  & Df & AIC & BIC & logLik & deviance & Chisq & Chi.Df & Pr..Chisq. \\
##   \hline
## object & 5.00 & 1668.76 & 1686.75 & -829.38 & 1658.76 &  &  &  \\
##   ..1 & 9.00 & 1670.18 & 1702.57 & -826.09 & 1652.18 & 6.58 & 4.00 & 0.16 \\
##    \hline
## \end{tabular}
## \end{table}
```