

Koffie, of: hoe vinden we niet bestaande relaties

Roel Hogervorst

31 oktober 2015

Intro

Op een mooie dag toen de wereld nog groen was hield Roel bij hoeveel koffie hij dronk. Ook Andreas hield koffie bij. We hielden zelfs bij hoeveel we samen dronken. Ik laat in meerdere stappen zien hoe koffie, het weer en de beurs samenhangen. Mijn data loopt van 2014-11-13 - 2015-6-5. In het ruwe rmarkdown bestand (rmd) zie je ook alle r code die gebruikt is om de grafieken te maken.

Benodigdheden: R versie 3.2 of hoger. (Dit is gemaakt in R version 3.2.2 (2015-08-14) – “Fire Safety”)
Packages gebruikt: readr, dplyr, ggplot2, cowplot, lubridate, GGally

Koffie

Databewerking en methode

De koffiedata was verzameld in losse .txt bestanden vanuit een log app die helaas over is gegaan naar de ‘darkside’ en waar je nu voor moet betalen(een betere optie voor android is nu tap log). Er zijn een aantal manuele handelingen vooraf gegaan: ik heb de logbestanden leeggehaald van andere log files en opgeslagen onder aparte namen want mijn logbestand bevatte ook de Samen data. De koffie samen en koffie Roel zijn nu een apart bestand geworden dat apart wordt ingeladen. Voor de weer en knmi data heb ik gebruik gemaakt van knmi.nl en yahoo finance, waarbij ik heb geselecteerd tussen de twee uiterste data van de koffie (2014-11-13 - 2015-6-5).

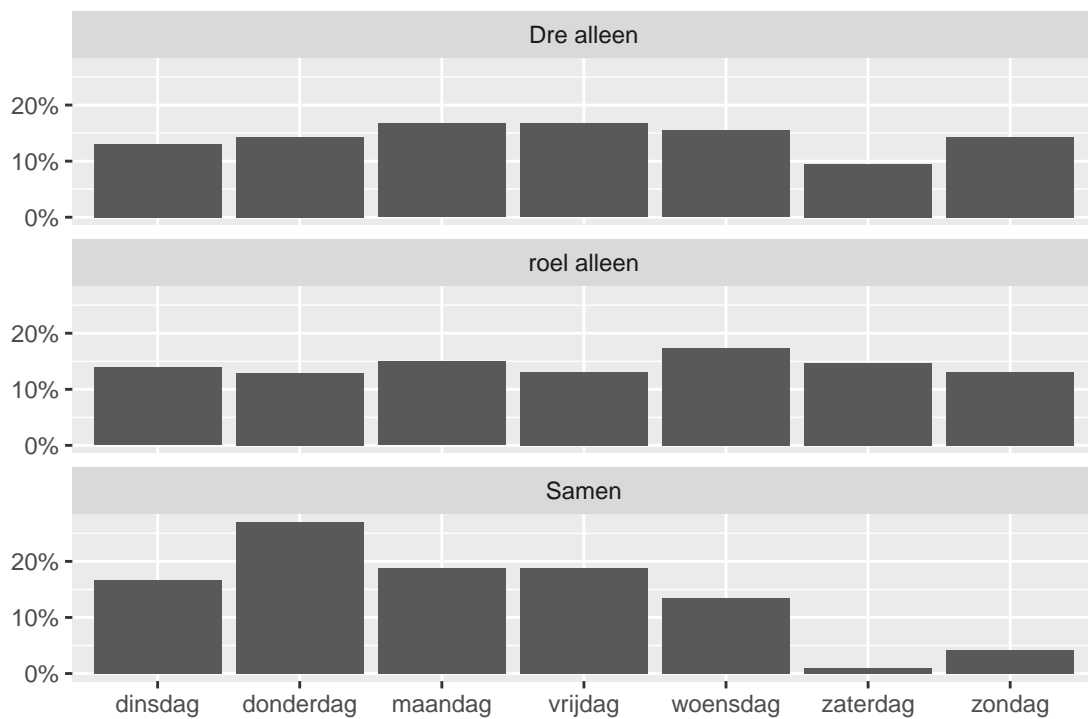
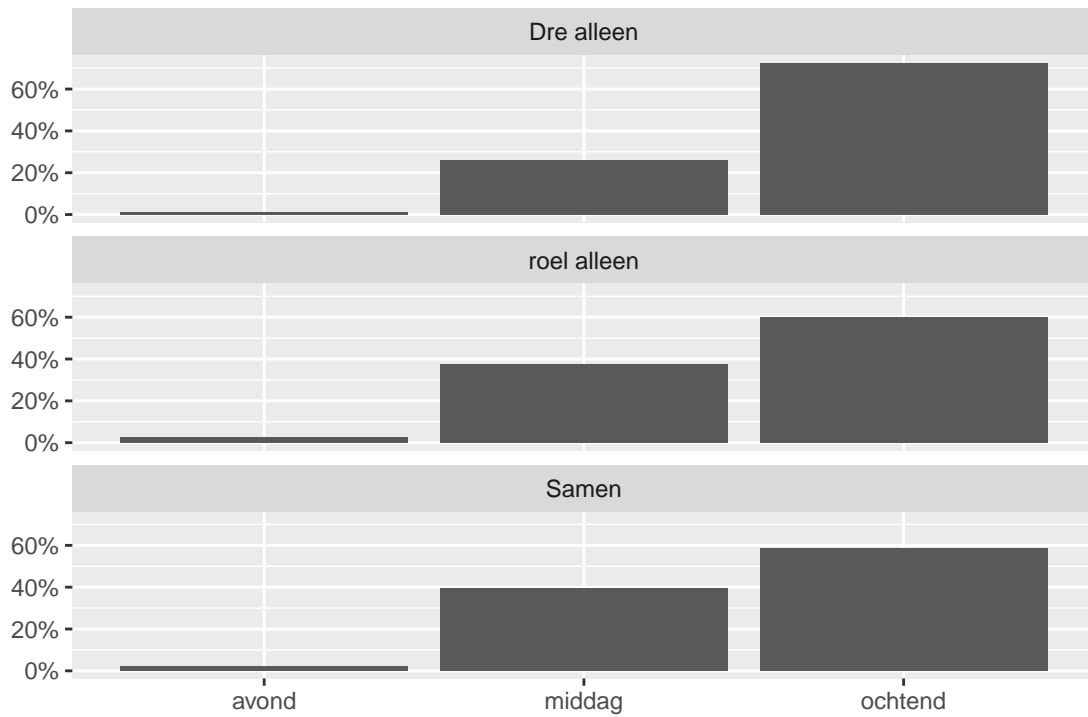
Eerst bespreek ik het koffiegebruik dan het weer en vervolgens de beurs (AEX). Hierna combineer ik de koffie, weer en beursdata.

korte visualisatie van koffie data

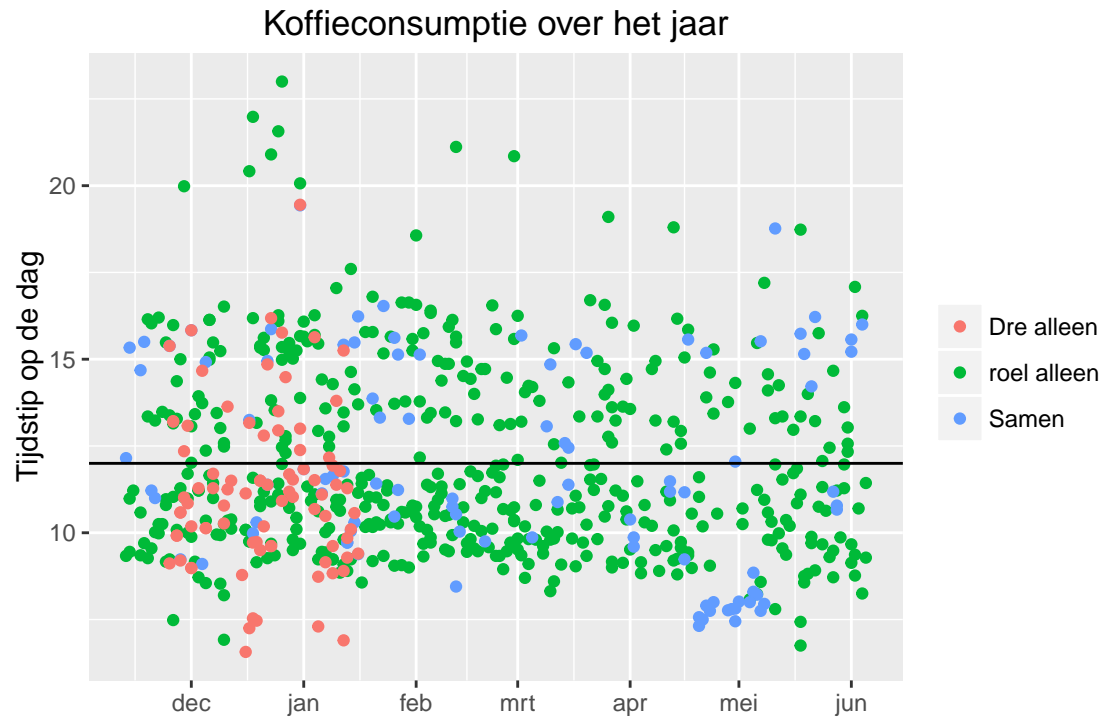
Dan visualiseren we de Coffee data in totaal. Ik gebruik ggplot en cowplot. De visualisatie van alles is niet heel zinnig omdat Roel veel meer koffie drinkt dan Andreas en Roel veel meer heeft bijgehouden dan Andreas zoals te zien is in de tabel.

##		avond	middag	ochtend
##	Dre alleen	1	22	61
##	roel alleen	13	180	289
##	Samen	2	38	56

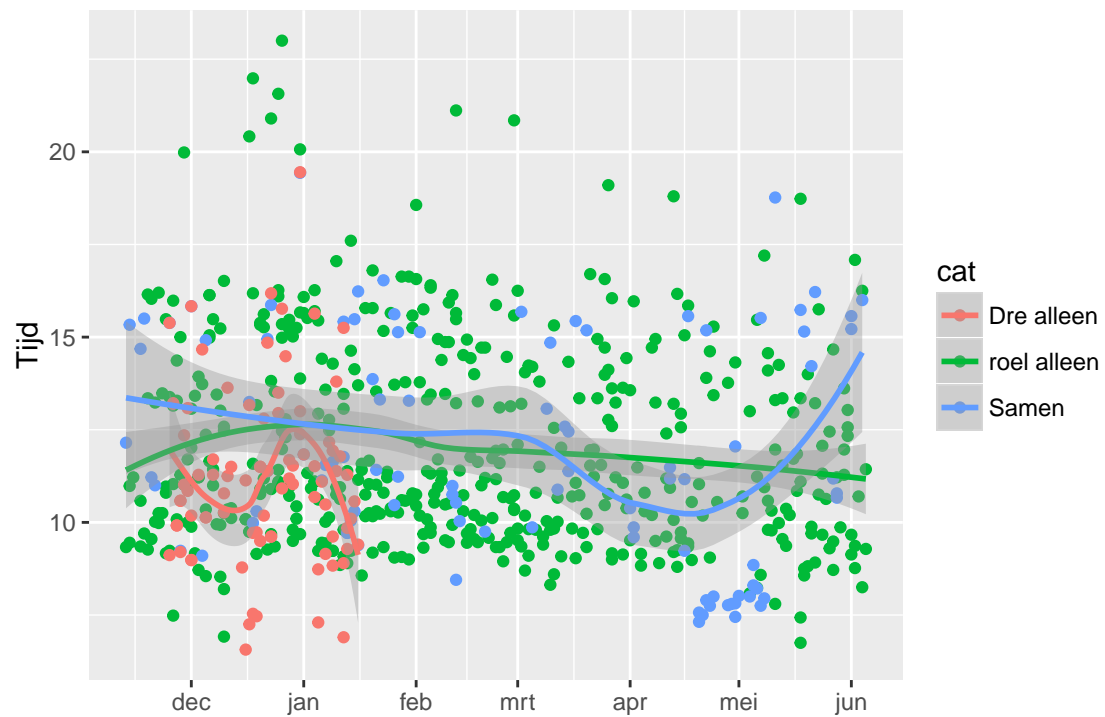
Je ziet hieronder hoe de koffie percentueel is verdeeld over de dag en over de week.

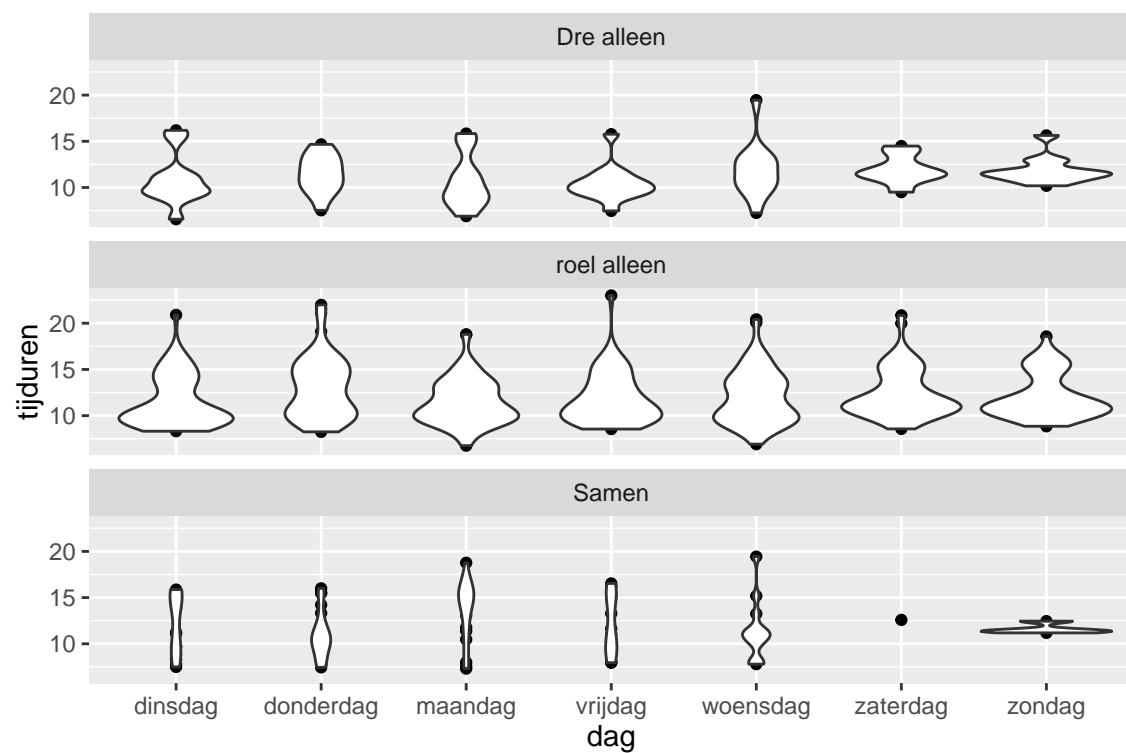
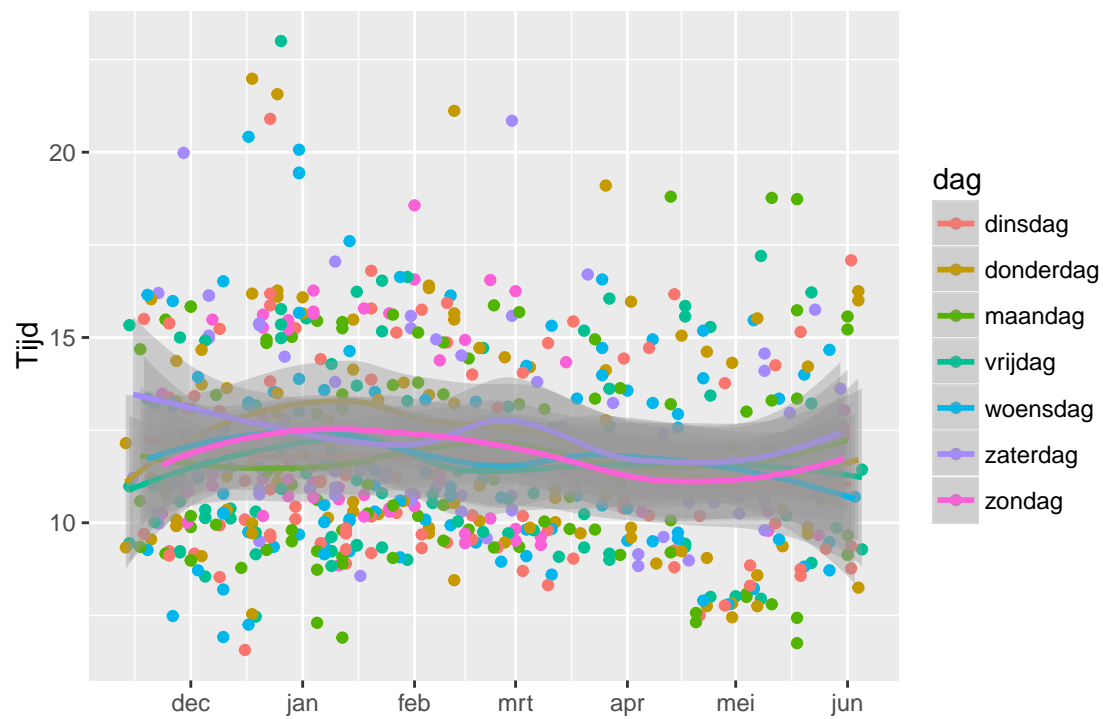


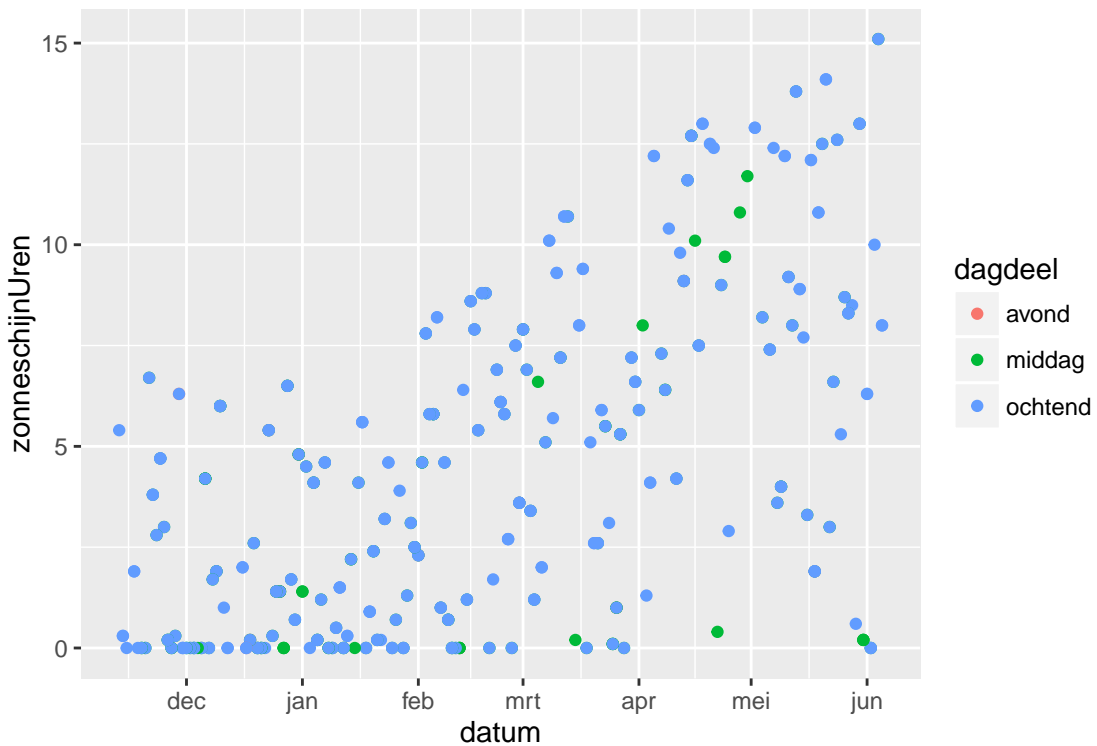
Vervolgens zijn de koffiemomenten per persoon over de tijd per dag geplot met een lijn op 12 uur (ik heb ook nog wat veranderd aan de as titels en de naam vande legenda weggehaald)



Laten we dan kijken naar patronen in koffiegebruik.

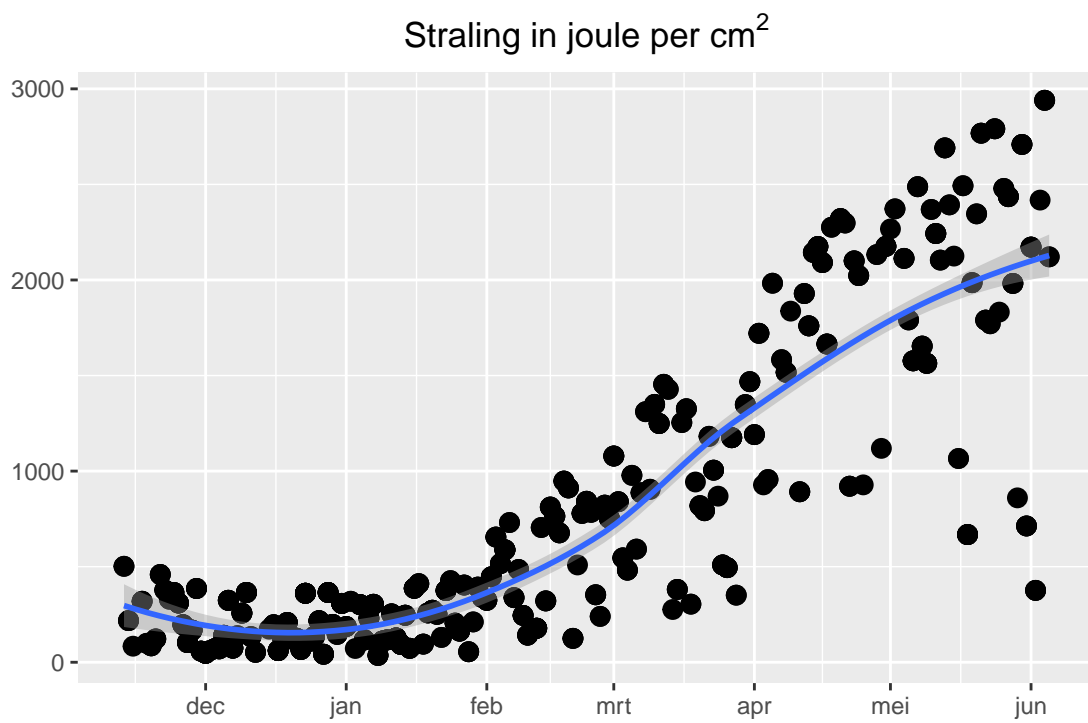
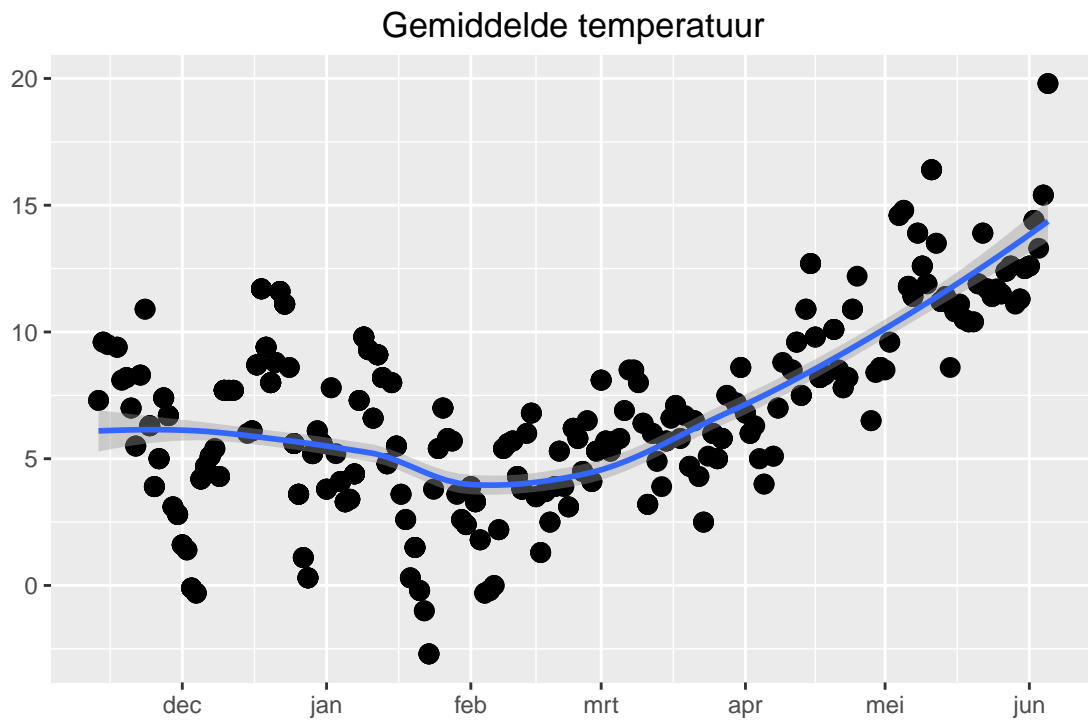


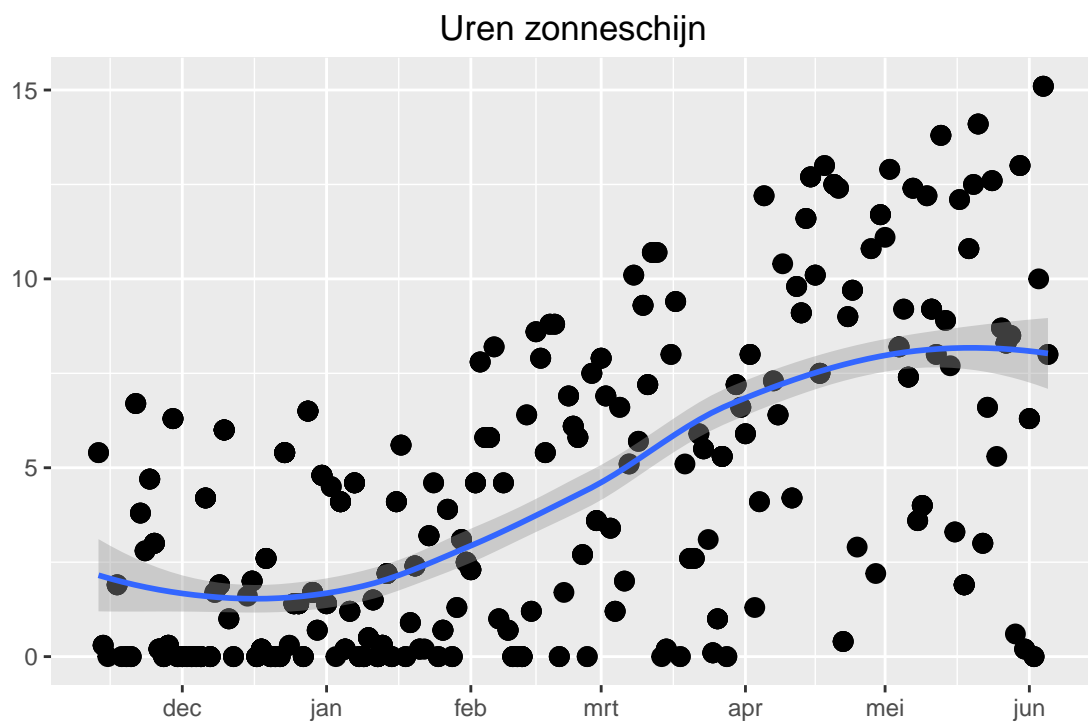
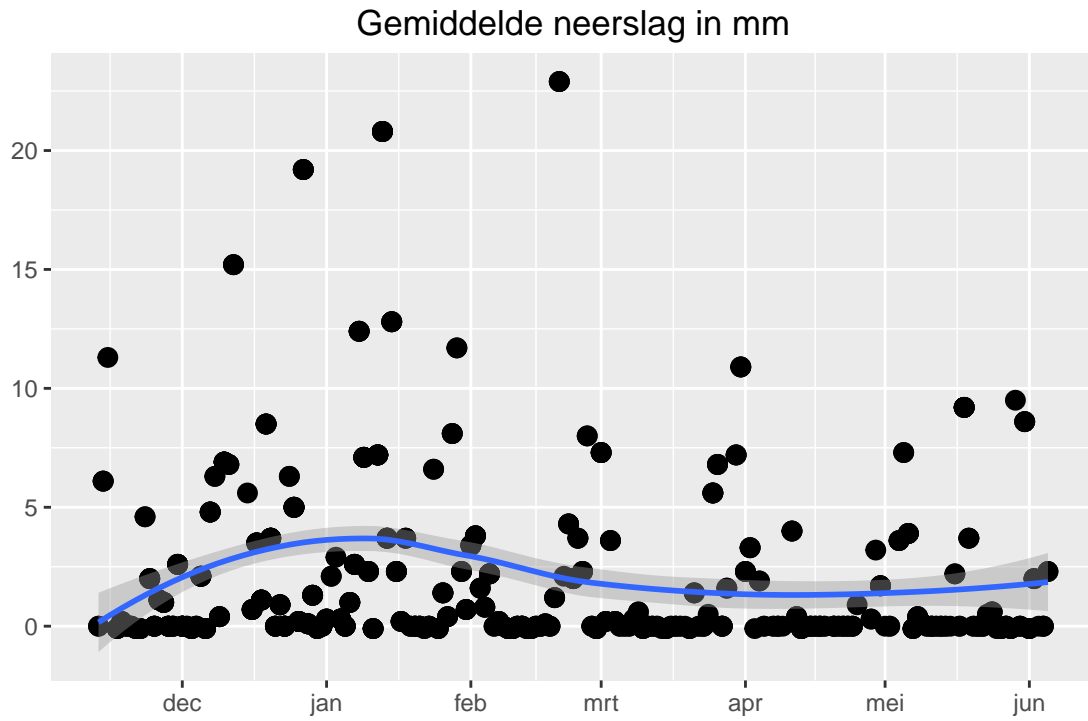


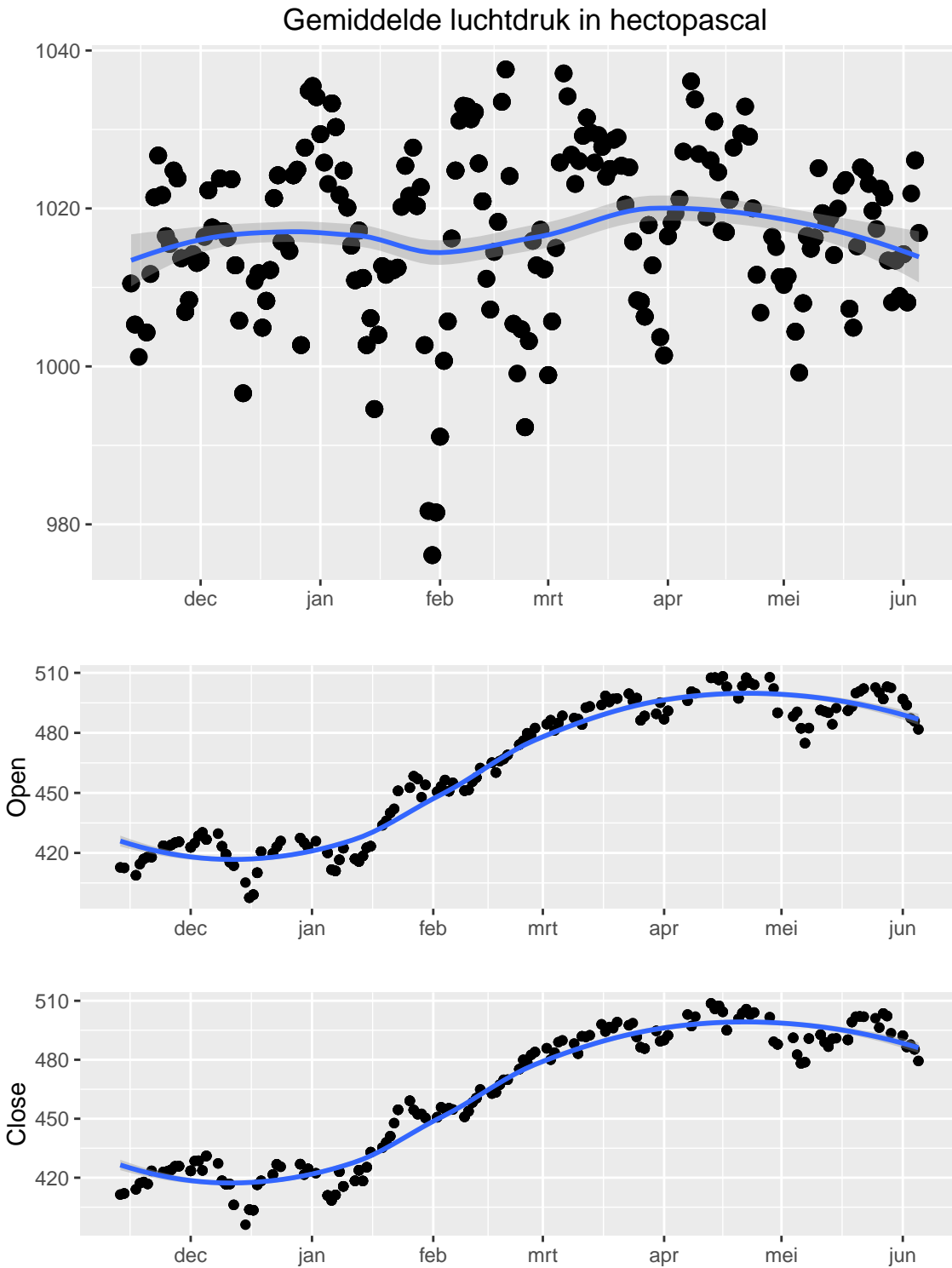


overzicht van weer en beurs data.

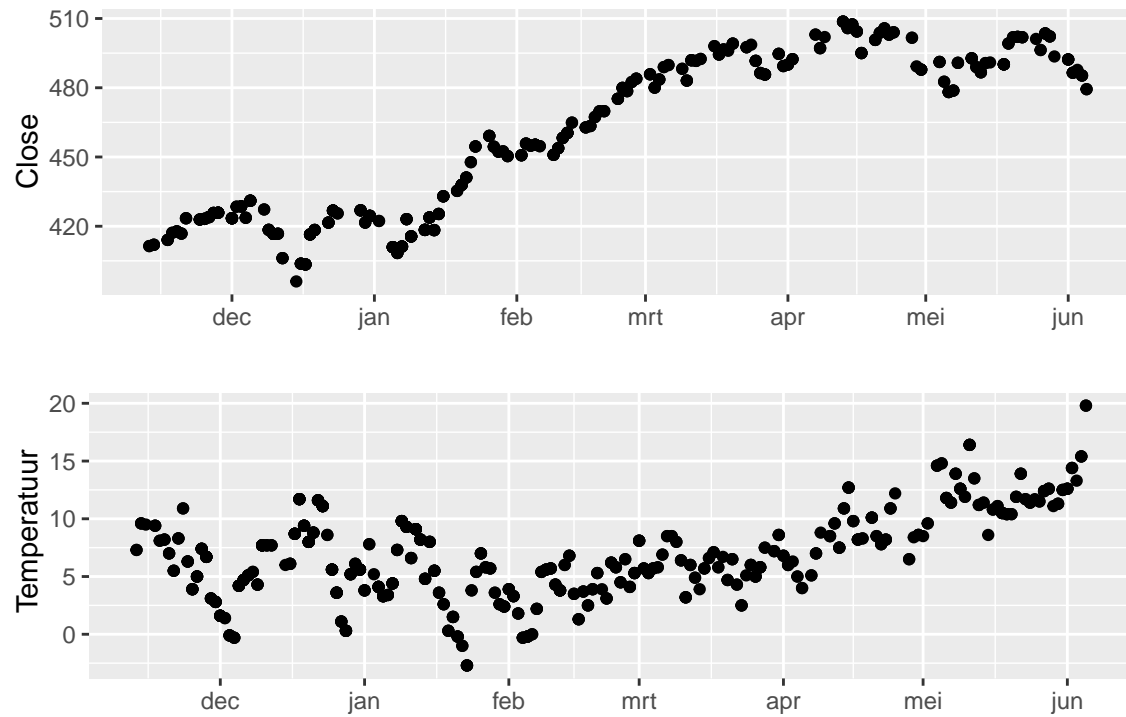
Eerst een overzicht van weer details over de gemeten periode en vervolgens een de opening en sluiting van de AEX. Helaas is de beurs gesloten op weekenden en dus mis ik daar een aantal waarden.







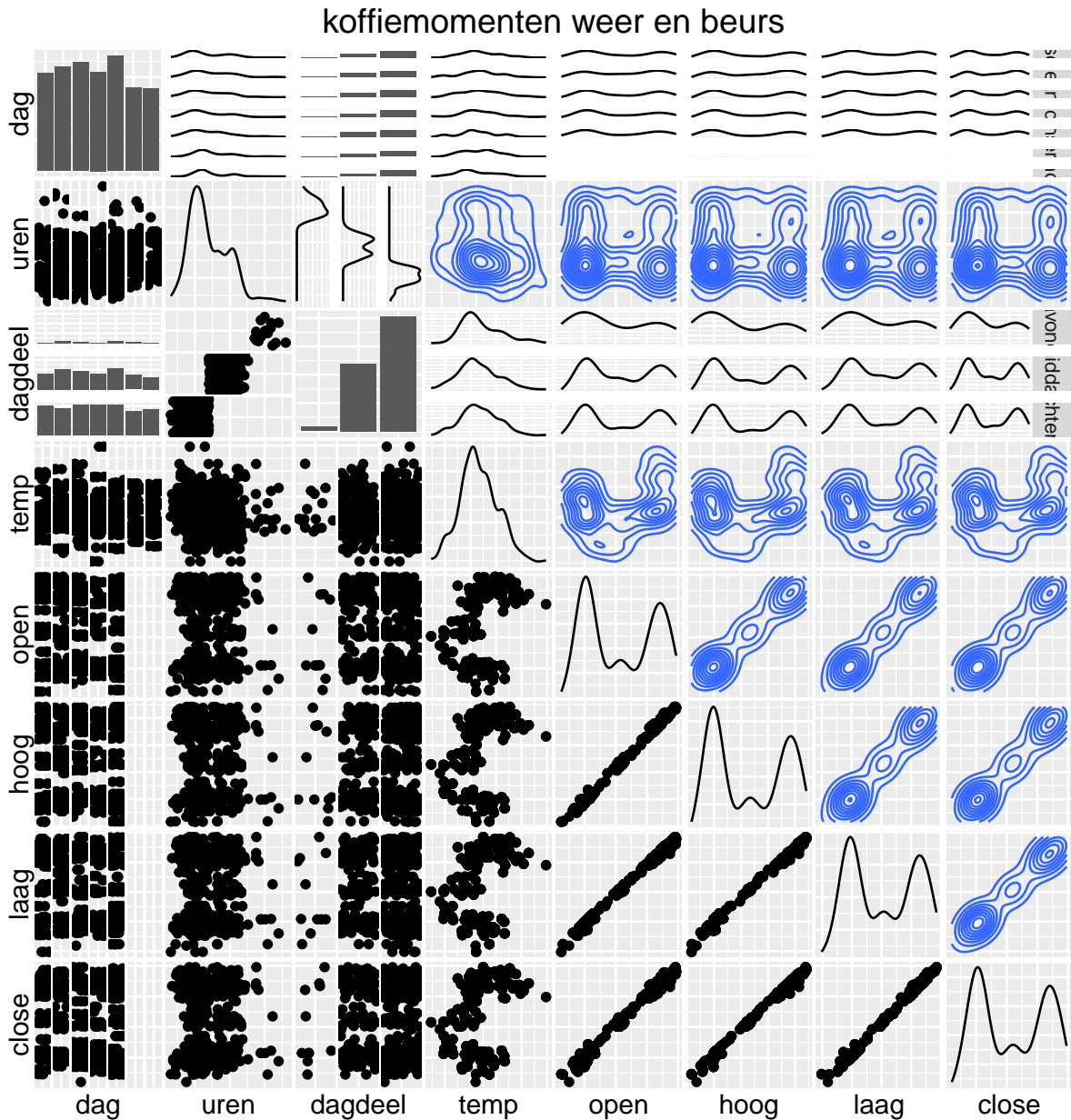
In de volgende visualisatie zie je de sluiting van de AEX en temperatuur. Ze zijn niet echt gerelateerd.



Analysen

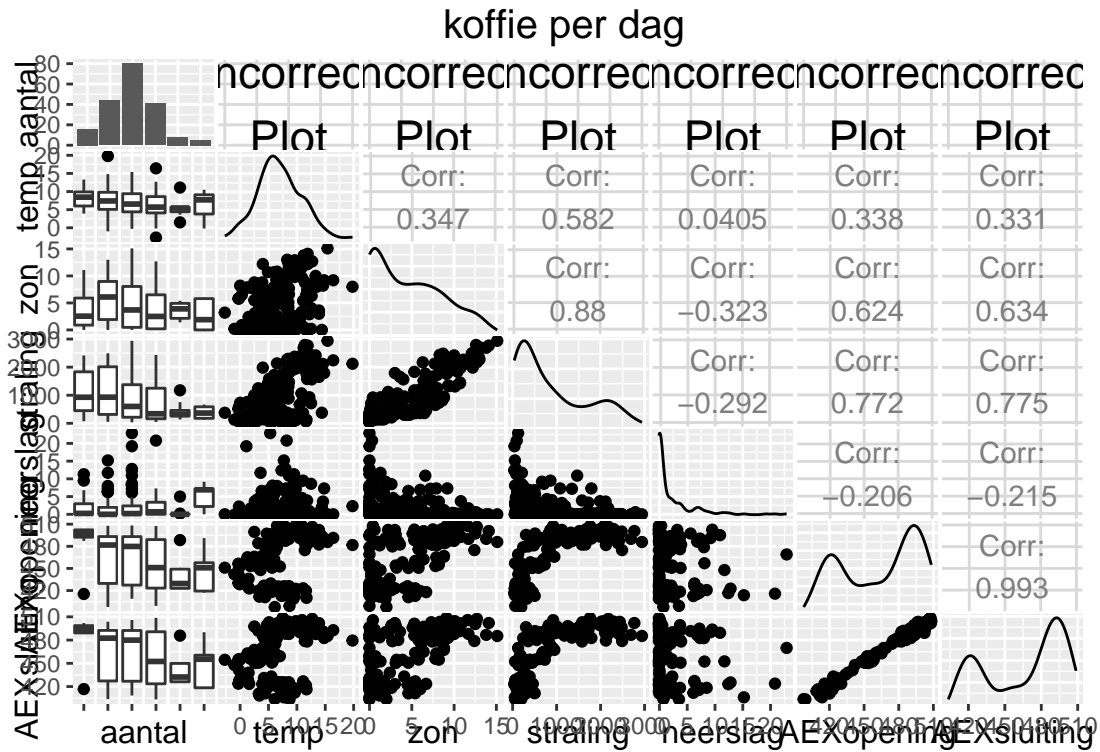
Correlaties tussen variabelen

Laten we dan eens kijken naar relaties tussen variabelen. Dat kan met een correlatie per paar, maar er is ook een coolere optie. In het volgende plot zie je alle relaties tussen verschillende variabelen. De kolommen en rijen zijn als volgt: dag (van de week), uren (tijdstip van koffie), dagdeel (ochtend middag, avond), temp (gemiddelde temperatuur op een dag), open (AEX van die dag opening).



Sommige verbanden zijn niet zo heel gek: de verdeling van uren over middag, avond, ochtend is logisch. De beurskoersen hebben geen informatie in de weekenden (dan is de beurs dicht). Denk eraan dat de informatie is gekoppeld aan koffie. Dus de temperaturen zijn veel meer in het ochtend gedeelte, maar dat komt omdat er vooral koffie wordt gedronken in de ochtend in dit dataset. Als er meerdere koppen koffie worden gedronken op een dag komen de temperatuur en beurskoersen van die dag dus meermaal terug.

Laten we daarom eens kijken naar een dataset geaggregeerd per dag.



Hier zien we de relaties tussen koffie beurs en weer per dag. Aantallen koffie lijken niet samen te hangen met andere waarden. Waar zien we grote correlaties: tussen temperatuur en straling .58 , tussen straling en zonne uren .88 , tussen AEX open en zon .77 maar ook tussen AEX sluit en zon .78 en natuurlijk tussen open en sluit koersen van de AEX .99.