

R Notebook - Kaan Gogcay

Tekst

```
library("tidyverse") # Tekst

## Registered S3 methods overwritten by 'ggplot2':
##   method           from
##   [.quosures       rlang
##   c.quosures       rlang
##   print.quosures  rlang

## Registered S3 method overwritten by 'rvest':
##   method           from
##   read_xml.response xml2

## — Attaching packages ————— tidyverse
1.2.1 —

## ✓ ggplot2 3.1.0      ✓ purrr   0.3.2
## ✓ tibble  2.1.1      ✓ dplyr   0.8.0.1
## ✓ tidyr   0.8.3      ✓ stringr 1.4.0
## ✓ readr   1.3.1      ✓ forcats 0.4.0

## — Conflicts —————
tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()

library("readxl")
```

Dataset_verkopen

```
excelddf <- read_xlsx("dataset_verkopen.xlsx")
excelddf

## # A tibble: 9 x 9
##   Naam  Geslacht Land  Continent Producten Opbrengst Dagen_geleden_1...
##   <chr> <chr>   <chr> <chr>      <chr>      <dbl>          <dbl>
## 1 Wim   m       Nede... Europa    iPhone      3000           5
## 2 John  m       Vere... Noord-Am... iPhone      2000          23
## 3 Ther... v       Fran... Europa    iPad        3450          15
## 4 Ling  v       China Azie    iPhone      1000         312
## 5 Ariko v       Japan Azie    iPad        2550         512
## 6 Fabr... m       Span... Europa    iPad        4100           8
## 7 Knut  m       Noor... Scandina... iPhone      6700          34
## 8 Ling  v       Taiw... Azie    iPad        3150         259
## 9 Mitch m       Cana... Noord-Am... iPhone      3500           8
## # ... with 2 more variables: Klanttevredenheid <dbl>, Aantal_besteld <dbl>
```

Werkmap

```
excelfdf <- read_xlsx("Werkmap.xlsx")
excelfdf

## # A tibble: 60 x 13
##   Gebruikers Leeftijd Geslacht Woonplaats Aantal_digitale... Abbonement
##   <chr>      <dbl> <chr>    <chr>          <dbl> <chr>
## 1 Jan Nonhof      73 M      Nederland      4 Free trial
## 2 Peter de ...    27 M      Nederland      2 Maandelij...
## 3 Joost van...    39 M      België        4 Maandelij...
## 4 Anahi Bus...    42 V      Verenigde...   4 Lifetime ...
## 5 Juan DeBi...    27 M      Argentinië     3 Maandelij...
## 6 Nina Rooy...    24 V      Nederland      5 Lifetime ...
## 7 Amber van...    25 V      Nederland      6 Free trial
## 8 Amish Bad...    32 M      Nederland      1 Lifetime ...
## 9 Coen Jans...    28 M      Verenigd ...   3 Maandelij...
## 10 Wouter Vo...   38 M      Nederland      2 Maandelij...
## # ... with 50 more rows, and 7 more variables: Extra_aankopen <chr>,
## #   Tijd_van_abbonee_in_dagen <dbl>,
## #   Hoevaak_apparaat_switchen_per_dag <dbl>, Inkomen_euro_per_maand <dbl>,
## #   Klanttevredenheid <dbl>, Lid_gebleven_na_free_trial <chr>,
## #   Hoe_bent_u_bij_ons_terecht_gekomen <chr>
```

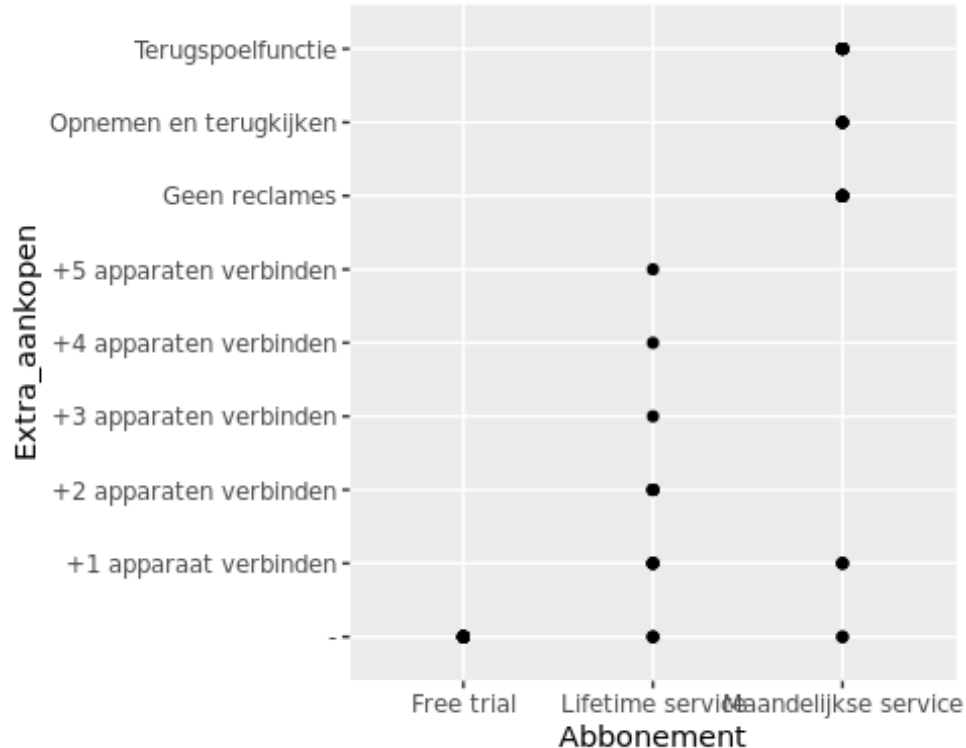
Mijn werkwijze

Welke extra's kopen mensen met hun abonnement?

```
excelfdf <- read_xlsx("Werkmap.xlsx")
excelfdf %>%
  arrange(desc(Hoe_bent_u_bij_ons_terecht_gekomen))

## # A tibble: 60 x 13
##   Gebruikers Leeftijd Geslacht Woonplaats Aantal_digitale... Abbonement
##   <chr>      <dbl> <chr>    <chr>          <dbl> <chr>
## 1 Nina Rooy...    24 V      Nederland      5 Lifetime ...
## 2 Rochan Ma...    23 M      Nederland      2 Maandelij...
## 3 Britt Fri...    19 V      Nederland      4 Free trial
## 4 Timo Den ...    20 M      Nederland      5 Maandelij...
## 5 Koen van ...    23 M      Nederland      3 Lifetime ...
## 6 Sven Smol...    23 M      Nederland      6 Maandelij...
## 7 Roy Huism...    20 M      Nederland      3 Maandelij...
## 8 Justin Ge...    17 M      Nederland      1 Lifetime ...
## 9 Jos Smull...    24 M      Nederland      4 Maandelij...
## 10 Jesse Cox      24 M      Nederland      4 Maandelij...
## # ... with 50 more rows, and 7 more variables: Extra_aankopen <chr>,
## #   Tijd_van_abbonee_in_dagen <dbl>,
## #   Hoevaak_apparaat_switchen_per_dag <dbl>, Inkomen_euro_per_maand <dbl>,
## #   Klanttevredenheid <dbl>, Lid_gebleven_na_free_trial <chr>,
## #   Hoe_bent_u_bij_ons_terecht_gekomen <chr>
```

```
ggplot(excelddf, aes(x = Abbonement, y = Extra_aankopen)) + geom_point()
```



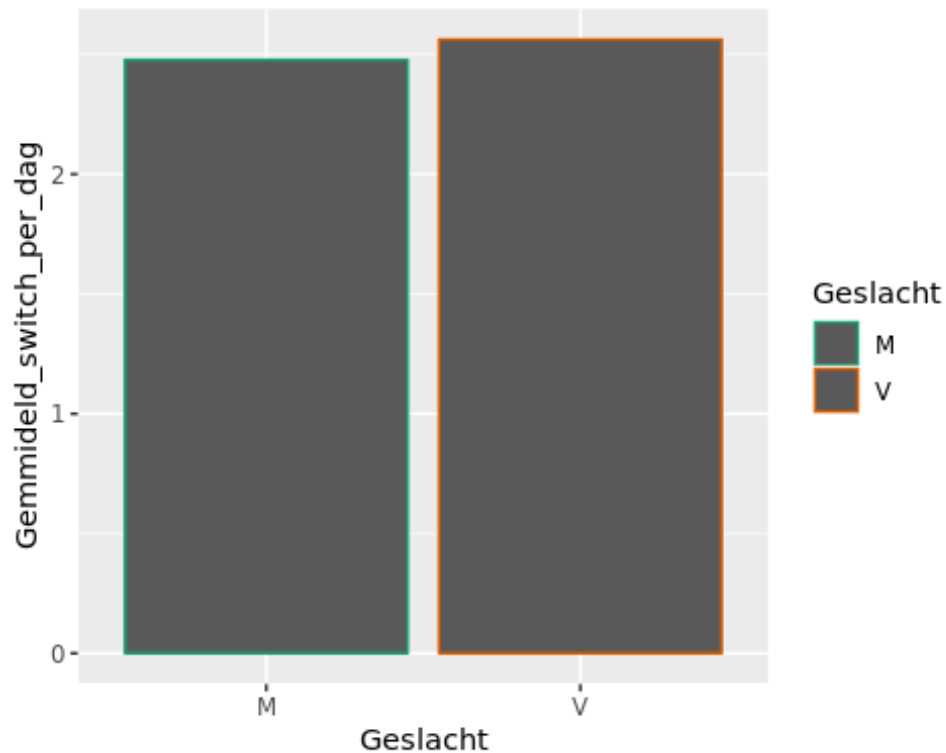
Conclusie: Mensen met de maandelijkse service willen niet meer dan 1 extra apparaat kopen. Bij de liftetime service wordt er wel flink gekocht.

#arrange, geom_point

Hoe vaak switchen de mensen gemmideld per dag, en verschilt dit tussen man en vrouw?

```
excelddf <- read_xlsx("Werkmap.xlsx")
by_geslacht <- excelddf%>%
  group_by(Geslacht)%>%
  summarize(Gemmideld_switch_per_dag = mean(Hoe vaak apparaat switchen per dag))

ggplot(by_geslacht, aes(x = Geslacht, y = Gemmideld_switch_per_dag, color =
  Geslacht)) + geom_col() + scale_color_brewer(palette="Dark2")
```



#Conclusie: De resultaten zijn er gelijkop. de vrouwen switchen per dag gemmiddeld wel ietjes meer dan de mannen.

#group_by, summarize, geom_col, scale_color_brewer,

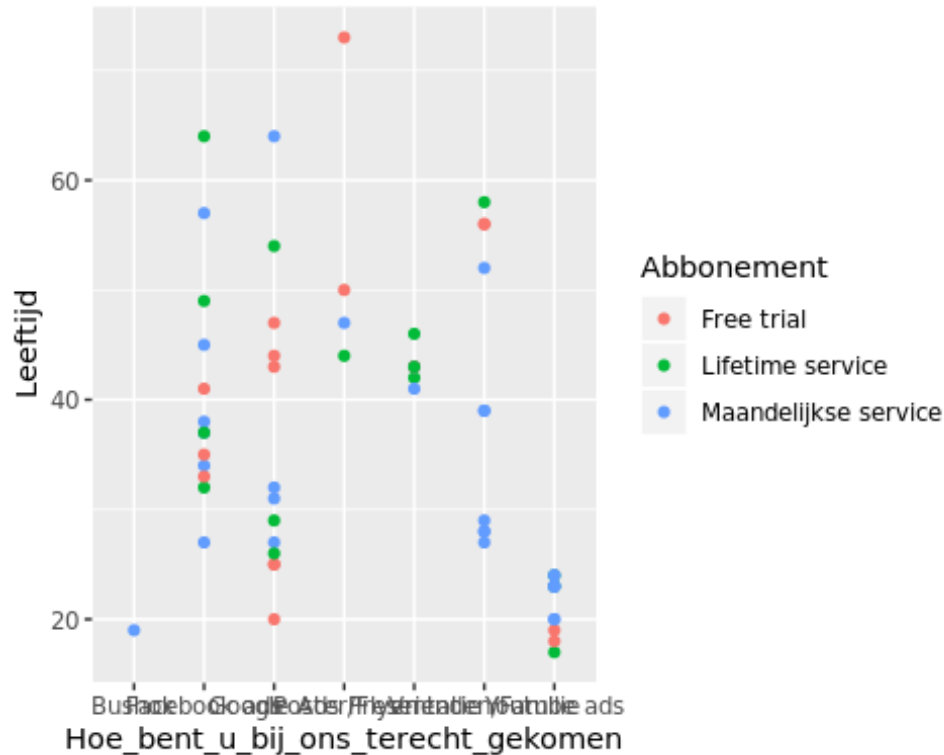
Hoe zijn ze bij ons terecht gekomen, en welk abonnement hebben ze dan gekozen?

```
excelddf <- read_xlsx("Werkmap.xlsx")
excelddf

## # A tibble: 60 x 13
##   Gebruikers Leeftijd Geslacht Woonplaats Aantal_digitale... Abbonement
##   <chr>      <dbl> <chr>    <chr>          <dbl> <chr>
## 1 Jan Nonhof      73 M      Nederland      4 Free trial
## 2 Peter de ...    27 M      Nederland      2 Maandelij...
## 3 Joost van...    39 M      België        4 Maandelij...
## 4 Anahi Bus...    42 V      Verenigde...  4 Lifetime ...
## 5 Juan DeBi...    27 M      Argentinië    3 Maandelij...
## 6 Nina Rooy...    24 V      Nederland     5 Lifetime ...
## 7 Amber van...    25 V      Nederland     6 Free trial
## 8 Amish Bad...    32 M      Nederland     1 Lifetime ...
## 9 Coen Jans...    28 M      Verenigd ...  3 Maandelij...
## 10 Wouter Vo...    38 M      Nederland     2 Maandelij...
## # ... with 50 more rows, and 7 more variables: Extra_aankopen <chr>,
## #   Tijd_van_abbonee_in_dagen <dbl>,
## #   Hoevaak_apparaat_switchen_per_dag <dbl>, Inkomen_euro_per_maand <dbl>,
```

```
## # Klanttevredenheid <dbl>, Lid_gebleven_na_free_trial <chr>,
## # Hoe_bent_u_bij_ons_terecht_gekomen <chr>

ggplot(excelddf, aes(x=Hoe_bent_u_bij_ons_terecht_gekomen, y=Leeftijd, color =
Abbonement)) + geom_point()
```



Conclusie: Mensen die van de presentatie komen kiezen bijna allemaal voor de Lifetime, dat komt denk ik omdat zij goed geïnformeerd zijn over de app. Ook zie ik veel Maandelijks bij vrienden/familie. En ik zie veel free trials bij Google ads. bij de andere bronnen kan ik niet echt conclusies trekken.

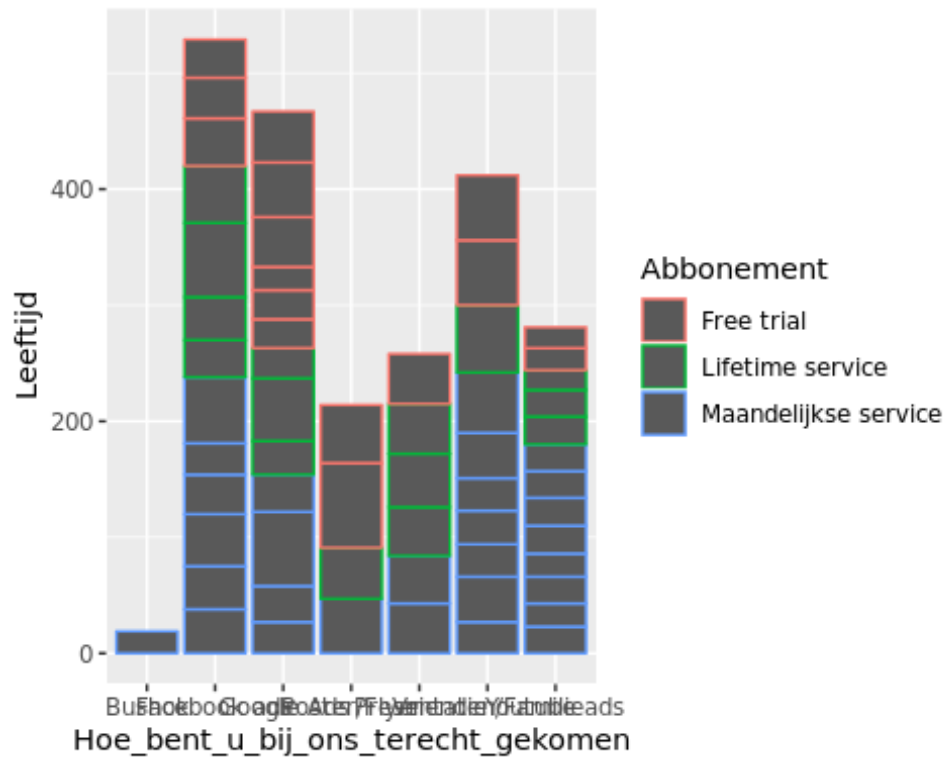
#Deze grafiek is niet zo duidelijk, hieronder een iets duidelijker grafiek.

#geom_point

Hoe zijn ze bij ons terecht gekomen, en welk abonnement hebben ze dan gekozen? (2)

```
excelddf <- read_xlsx("Werkmap.xlsx")
by_Hoe_bent_u_bij_ons_terecht_gekomen <- excelddf %>%
  group_by(Hoe_bent_u_bij_ons_terecht_gekomen)

ggplot(excelddf, aes(x=Hoe_bent_u_bij_ons_terecht_gekomen, y=Leeftijd, color =
Abbonement)) + geom_col()
```



Conclusie: Mensen die van de poster komen gaan eerder voor de free trial. Ook zie je de free trial nog veel bij de google ads. Voor de rest is het vooral Maandelijks en een beetje Lifetime.

#group_by, geom_col

Filter op 25 jaar en ouder, Leeftijd van hoog naar laag.

```
exceldf <- read_xlsx("Werkmap.xlsx")
exceldf %>%
  filter(Leeftijd >= 25)%>%
  arrange(desc(Leeftijd))
```

A tibble: 45 x 13

##	Gebruikers	Leeftijd	Geslacht	Woonplaats	Aantal_digitale...	Abbonement
##	<chr>	<dbl>	<chr>	<chr>	<dbl>	<chr>
## 1	Jan Nonhof	73	M	Nederland	4	Free trial
## 2	Elia van...	64	M	Nederland	2	Lifetime ...
## 3	Gijs Vijf...	64	M	Nederland	3	Maandelij...
## 4	Flora Sta...	58	V	Verenigde...	5	Lifetime ...
## 5	Nancy Van...	57	V	Nederland	4	Maandelij...
## 6	Diana Alm...	56	V	Nederland	4	Free trial
## 7	Arnold Mo...	56	M	Verenigde...	3	Free trial
## 8	Joshua do...	54	M	Nederland	4	Lifetime ...
## 9	Gijs Vang...	52	M	Nederland	5	Maandelij...
## 10	Alex Romb...	50	M	Frankrijk	3	Free trial

... with 35 more rows, and 7 more variables: Extra_aankopen <chr>,

```
## # Tijd_van_abbonee_in_dagen <dbl>,
## # Hoevaak_apparaat_switchen_per_dag <dbl>, Inkomen_euro_per_maand <dbl>,
## # Klanttevredenheid <dbl>, Lid_gebleven_na_free_trial <chr>,
## # Hoe_bent_u_bij_ons_terecht_gekomen <chr>
```

#Conclusie: We hebben best veel klanten boven de 25.

#filter, arrange

Mutate, Hoeveel verdienen onze klanten per jaar en per dag

```
exceldf <- read_xlsx("Werkmap.xlsx")
exceldf %>%
  mutate(Inkomen_euro_per_jaar = Inkomen_euro_per_maand * 12,
  Inkomen_euro_per_dag = Inkomen_euro_per_maand / 30)

## # A tibble: 60 x 15
##   Gebruikers Leeftijd Geslacht Woonplaats Aantal_digitale... Abbonement
##   <chr>         <dbl> <chr>    <chr>          <dbl> <chr>
## 1 Jan Nonhof      73 M      Nederland          4 Free trial
## 2 Peter de ...    27 M      Nederland          2 Maandelij...
## 3 Joost van...    39 M      België            4 Maandelij...
## 4 Anahi Bus...    42 V      Verenigde...      4 Lifetime ...
## 5 Juan DeBi...    27 M      Argentinië        3 Maandelij...
## 6 Nina Rooy...    24 V      Nederland          5 Lifetime ...
## 7 Amber van...    25 V      Nederland          6 Free trial
## 8 Amish Bad...    32 M      Nederland          1 Lifetime ...
## 9 Coen Jans...    28 M      Verenigd ...      3 Maandelij...
## 10 Wouter Vo...   38 M      Nederland          2 Maandelij...
## # ... with 50 more rows, and 9 more variables: Extra_aankopen <chr>,
## #   Tijd_van_abbonee_in_dagen <dbl>,
## #   Hoevaak_apparaat_switchen_per_dag <dbl>, Inkomen_euro_per_maand <dbl>,
## #   Klanttevredenheid <dbl>, Lid_gebleven_na_free_trial <chr>,
## #   Hoe_bent_u_bij_ons_terecht_gekomen <chr>, Inkomen_euro_per_jaar <dbl>,
## #   Inkomen_euro_per_dag <dbl>
```

#mutate

Gemiddelde Leeftijd en Totaal aantal switches per dag

```
exceldf <- read_xlsx("Werkmap.xlsx")
exceldf %>%
  group_by(Geslacht) %>%
  summarize(Gemiddelde_Leeftijd = median(Leeftijd),
  Totaal_aantal_swicthes_per_dag = sum(Hoevaak_apparaat_switchen_per_dag))

## # A tibble: 2 x 3
##   Geslacht Gemiddelde_Leeftijd Totaal_aantal_swicthes_per_dag
##   <chr>         <dbl>          <dbl>
## 1 M              34.5             109
## 2 V              35              41
```

#Conclusie: Gemiddelde leeftijd zit rond 35, het totaal aantal switches bij mannen is veel groter. Dat komt omdat we meer mannelijke gebruikers hebben.

#group_by, summarize

Wat zijn de gemiddelde inkomen van onze klanten en verschilt dit tussen man en vrouw?

```
exceldf <- read_xlsx("Werkmap.xlsx")
exceldf %>%
  group_by(Geslacht) %>%
  summarize(Gemmiddeld_inkomen = mean(Inkomen_euro_per_maand))
```

```
## # A tibble: 2 x 2
##   Geslacht Gemmiddeld_inkomen
##   <chr>          <dbl>
## 1 M             1716.
## 2 V             1847.
```

#Conclusie: Vrouwen verdienen gemiddeld meer dan mannen.

#group_by, summarize

Advies Bij de maandelijkse service zijn mensen niet zo snel getriggerd om extra's te kopen. ik stel voor dat we de maandelijkse service iets meer gaan limitten. Normale functies die er normaal inzitten uithalen en mensen voor laten betalen. de andere services staan nu op een goede positie.

Ook is het belangrijk dat we duidelijker gaan adverteren. De presentatie waar we onze product hebben gepresenteerd heeft veel mensen gemotiveerd de lifetime service te kopen. Maar bij de andere advertenties is het vooral de maandelijkse service die gekocht wordt. Daar kunnen we verandering in brengen. We moeten de posters en de advertenties op google duidelijk maken en mensen beter overtuigen toch voor de lifetime service te gaan.

Het gemiddelde leeftijd zit rond de 35 dat is niet echt een verrassing. We moeten proberen om ook jongere aan te spreken met onze software. Jongeren zitten vooral op youtube en netflix. Het lijkt me een goed idee om oude films te zetten op onze software. dat je naast live televisie kijken ook oude films kunt kijken.

We hebben gemerkt dat we veel meer mannelijke gebruikers hebben. Ik weet niet of het mogelijk is, maar misschien een korting voor vrouwen. zo kunnen we meer vrouwelijke gebruikers krijgen.