

Onderzoek Grafieken in R

Created by Lars van Duijnhoven, last modified on Dec 15, 2023

- [Inleiding](#)
- [Belang van het onderzoek](#)
- [Onderzoeksvragen](#)
 - [Hoofdvraag:](#)
 - [Deelvragen:](#)
- [Onderzoeksmethodiek](#)
- [Criteria en rationale](#)
- [Oplosrichtingen](#)
- [Resultaten](#)
 - [Deelvragen:](#)
 - [Welke grafieken worden regelmatig gebruikt voor een simulatie en hebben deze nut voor ons?](#)
 - [Van een lijst van 10 grafieksoorten die ondersteunt worden door R, welke zijn voor ons het beste om te gebruiken en waarom? Hiervan moeten er minimaal 3 verder uitgewerkt worden.](#)
 - [Bar graph \(Holtz, z.d.-b\)](#)
 - [Donut chart \(Holtz, z.d.-e\)](#)
 - [Line chart \(Holtz, z.d.-g\)](#)
 - [Pie-chart \(Holtz, z.d.-h\)](#)
 - [Treemap \(Holtz, z.d.-i\)](#)
 - [Violin chart \(Holtz, z.d.-j\)](#)
 - [Histogram \(Holtz, z.d.-f\)](#)
 - [Circle packing \(Holtz, z.d.-c\)](#)
 - [Density 2d \(Holtz, z.d.-a\)](#)
 - [Density plot \(Holtz, z.d.-d\)](#)
 - [Hoofdvraag:](#)
 - [Welke grafieken gaan wij gebruiken voor ons rapport in R?](#)
 - [Bar graph \(Holtz, z.d.-b\)](#)
 - [Donut chart \(Holtz, z.d.-e\)](#)
 - [Line chart \(Holtz, z.d.-g\)](#)
 - [Violin chart \(Holtz, z.d.-j\)](#)
- [Conclusie](#)
- [Bibliografie](#)

Inleiding

Om ervoor te zorgen dat wij onze informatie op een informatieve en nuttige manier weergeven moet onze gegenereerde informatie in een juiste grafiek getoond worden.

Dit onderzoek is bedoeld om te zorgen dat wij een rapport opleveren waar de toekomstige medewerkers echt wat aan hebben, zodat ze zo snel en secuur mogelijk een conclusie kunnen trekken uit de cijfertjes.

Belang van het onderzoek

Dit onderzoek is belangrijk om ervoor te zorgen dat de simulatiegegevens op een correcte manier weergegeven worden. Hiermee wordt bedoeld dat het niet op een misleidende manier wordt weergegeven, waardoor de lezer een verkeerde gedachte erover krijgt. Daarnaast zijn er ook grafieken die bepaalde veranderingen beter kunnen weergeven dan andere. Deze Sprint (Sprint 3) wordt het rapport gegenereerd in R en dus is het handig om dit onderzoek nu uit te voeren. Dit betekent ook dat het onderzoek vóór het rapport genereren afgerond moet zijn, wat de noodzaak dus verhoogd. Hierom hebben we gekozen om dit onderzoek binnen de eerste week van Sprint 3 uit te voeren.

Onderzoeksvragen

Hoofdvraag:

Welke grafieken gaan wij gebruiken voor ons rapport in R?

Deelvragen:

Welke grafieken worden het meest gebruikt voor een onderzoek om fabrieksinformatie/ simulatiegegevens te ewergeven en hebben deze nut voor ons project?

Van een lijst van 10 grafieksoorten die ondersteunt worden door R, welke zijn voor ons het beste om te gebruiken en waarom? Hiervan moeten er minimaal 3 verder uitgewerkt worden.

Onderzoeksmethodiek

Als onderzoeksmethode heb ik gekozen om een Literature Study ([Vogel, z.d.-a](#)) te gaan uitvoeren. Dit lijkt mij de beste manier om te onderzoeken, omdat er op internet waarschijnlijk genoeg informatie staat over grafieken binnen R en welke grafieken in de fabriekswereld veel gebruikt worden.

Criteria en rationale

Ik ga de uiteindelijke grafieken beoordelen op 3 gebieden, namelijk:

- **Overzichtelijkheid**

De grafieken moeten overzichtelijk zijn en er moet gelijk duidelijk zijn wat er staat. Dit houdt in dat je niet 300x moet inzoomen om tot nuttige informatie te komen, maar dat je dit het liefste gelijk al ziet. Dit houdt dus ook in dat de grafieken niet te groot mogen worden.

- **Dekking van lading**

De grafieken moeten de lading dekken, hiermee wordt bedoeld dat de grafiek helpt met het krijgen van een idee van wat er in de simulatie omgaat. De grafiek moet niet meer vragen oproepen, maar juist vragen beantwoorden en een inzicht geven in wat er in de simulatie is omgegaan.

- **Misleidbaarheid**

De grafieken moeten niet misleidend zijn. Hiermee bedoelen we dat er zo min mogelijk eigen invulling voor nodig is en dat er geen verkeerde conclusies uit getrokken kunnen worden, want dit zou grote gevolgen kunnen hebben. Als een grafiek dus slecht te lezen is, wordt het onduidelijk en kan de lezer eigen aannames maken.

Oplosrichtingen

Zoals eerder verteld in 1.4 ga ik een Literature Study uitvoeren door informatie op het internet op te zoeken. Hiervoor heb ik meerdere sites gebruikt, namelijk ([Holtz, z.d.-i](#)), ([Rindhe, 2021](#)) en ([Oetting, 2023](#)). Van deze sites heb ik informatie opgedaan over de verschillende soorten grafieken, wat de doelen hiervan zijn (Zie bronnenlijst voor alle specifieke links), welke grafieken het meeste gebruikt worden, wat de voor- en nadelen zijn van bijvoorbeeld een pie-chart e.t.c.

Resultaten

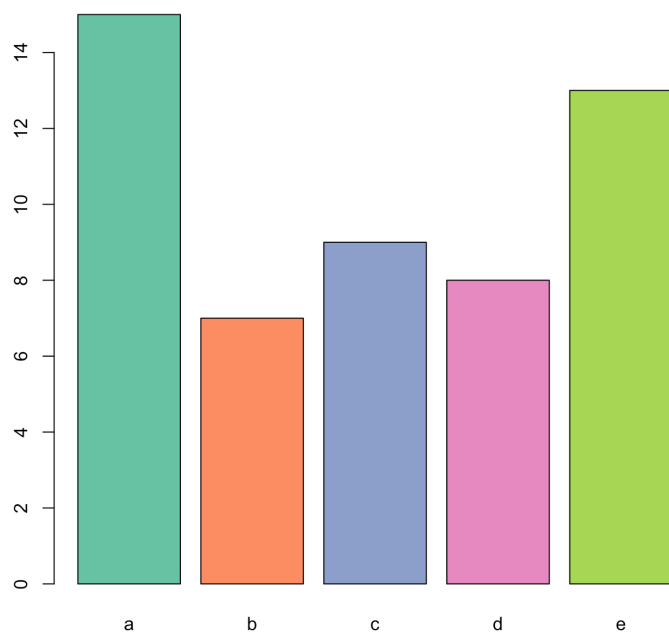
Deelvragen:

Welke grafieken worden regelmatig gebruikt voor een simulatie en hebben deze nut voor ons?

Hiervoor heb ik 2 bronnen gebruikt, namelijk ([Rindhe, 2021](#)) en ([Oetting, 2023](#)). De grafieken die in beide websites te vinden zijn, zijn de Bar Chart, Histogram, Pie-Chart, Donut Chart en Treemap Chart. Deze grafieken worden allemaal ondersteunt in R en zal ik dus hieronder in de andere deelvraag verder uitwerken.

Van een lijst van 10 grafieksoorten die ondersteunt worden door R, welke zijn voor ons het beste om te gebruiken en waarom? Hiervan moeten er minimaal 3 verder uitgewerkt worden.

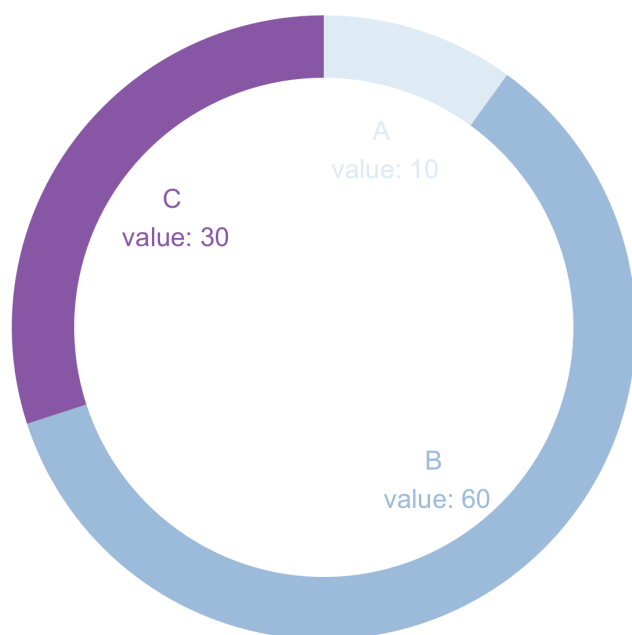
Bar graph ([Holtz, z.d.-b](#))



"A barplot is used to display the relationship between a numeric and a categorical variable".

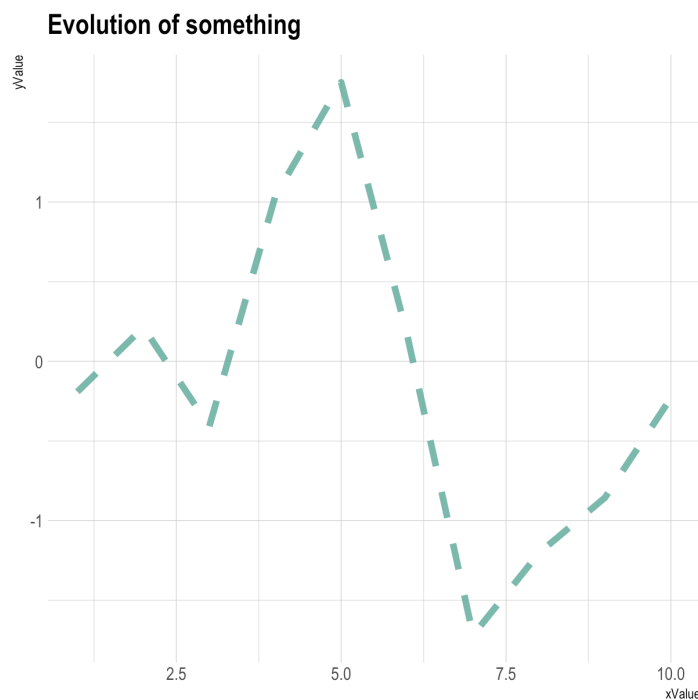
Het doel van deze grafiek is dus om een relatie te laten zien tussen bijvoorbeeld een machine en het aantal geproduceerde producten. Dit kan best handig zijn voor ons rapport, omdat per machine erg makkelijk de relatie met andere machines te zien is. Daarnaast stond was dit een van de grafieken die veel gebruikt werd binnen het fabrieksgebeid, waardoor dit al een van de grafieken is die verder uitgewerkt worden.

Donut chart (Holtz, z.d.-e)



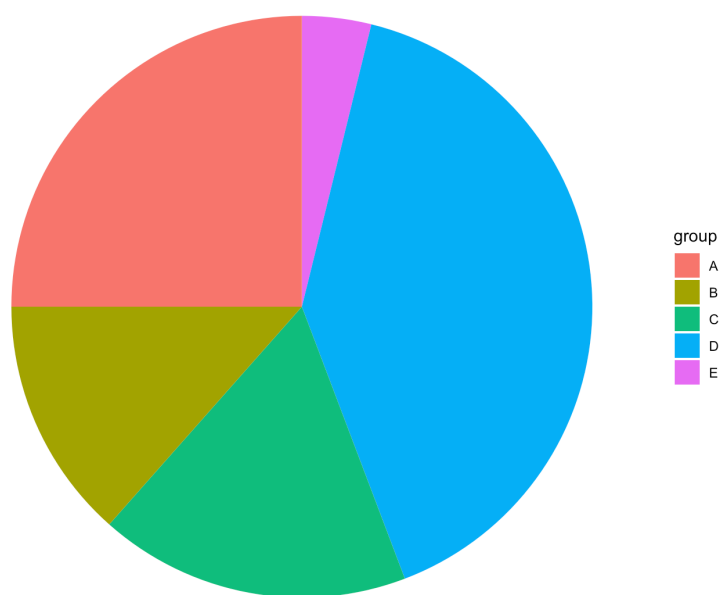
Donut charts zijn wel handig om een klein aantal grote waarden weer te geven, maar deze waarden mogen niet te veel verschillen van elkaar, want dan wordt het onoverzichtelijk. Daarnaast is er ook nog een probleem, voor de uitleg hiervan zie het kopje Pie Chart. Ondanks deze 2 problemen is er alsnog gekozen om verder in te gaan op de Donut chart, omdat we al een idee hebben van wat we ermee gaan visualiseren, namelijk de up- en downtime.

Line chart (Holtz, z.d.-g)



Iedereen kent hem wel, de standaard grafiek, namelijk de Line Chart. Dit is een duidelijke grafiek die op veel verschillende manieren te gebruiken is en in het algemeen een van de meest gebruikte grafieken is (Oetting, 2023). We gaan dus wel verder in op de Line Graph.

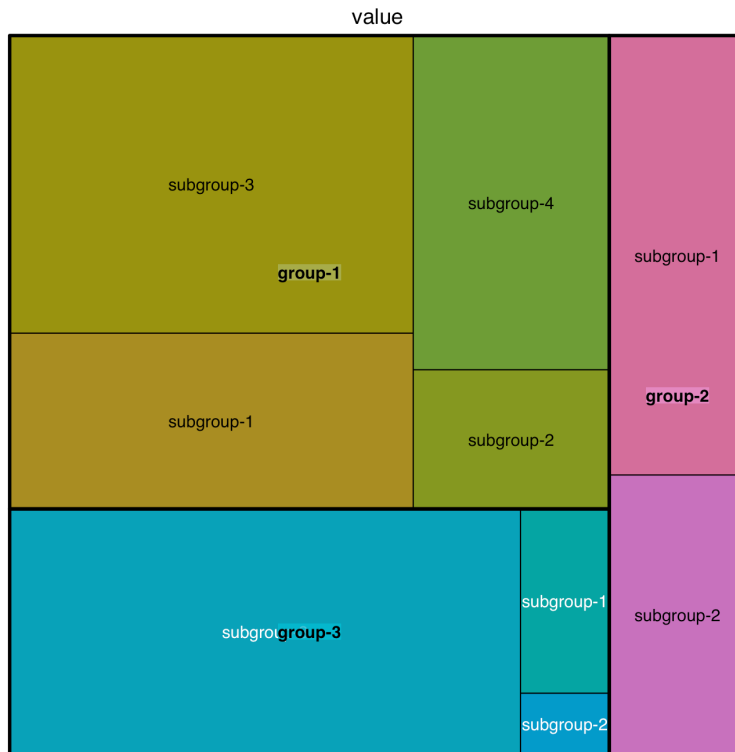
Pie-chart (Holtz, z.d.-h)



Zoals bij de Donut chart al gezegd is, heeft de Pie-chart een erg negatieve eigenschap (die de Donut chart dus ook

heeft). Dit is namelijk dat het moeilijk is om veranderingen of verschillen weer te geven in zo'n grafiek, omdat dat nou eenmaal is hoe het gebouwd is. (*Pie-chart bad by definition, z.d.*) Helaas vonden wij, ondanks het feit dat deze in de fabriekswereld veelgebruikt is ([Rindhe, 2021](#)), de Pie-chart niet nuttig genoeg om verder op in te gaan, mede dankzij deze eigenschap die het gebruik van meerdere diagrammen min of meer blokkeert.

Treemap ([Holtz, z.d.-i](#))



"A Treemap displays hierarchical data as a set of nested rectangles. Each group is represented by a rectangle, which area is proportional to its value."

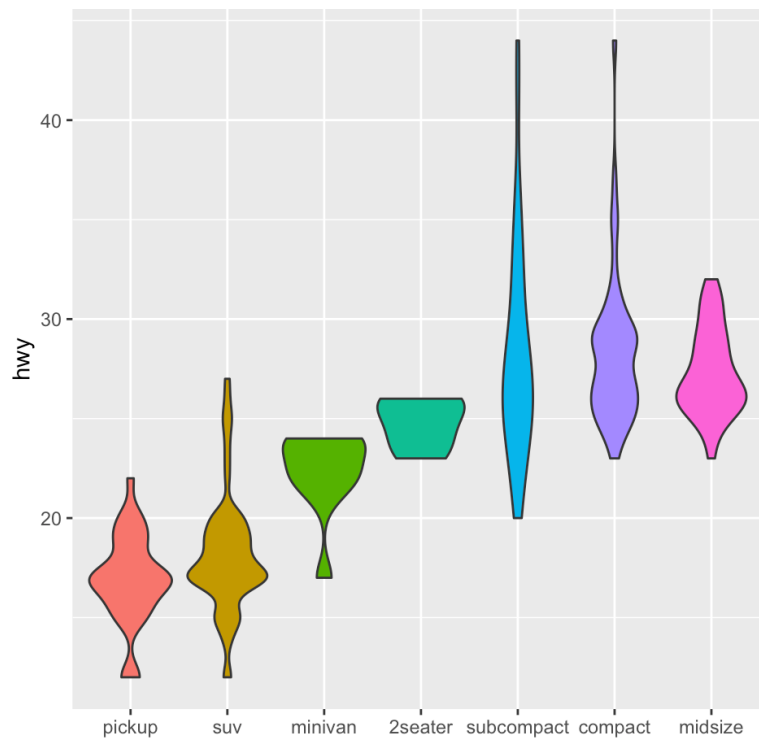
Een Treemap geeft een soort hiërarchie weer van grote rechthoeken die dan opgesplitst kunnen worden in andere kleinere subgroepen, zoals bij de afbeelding te zien is.

Deze groepen zijn gescheiden met kleur en lijnen, waardoor de hiërarchie zichtbaar is binnen het diagram.

Daarnaast is de Treemap een veelgebruikt diagram bij het weergeven van informatie binnen de fabriekswereld. ([Rindhe, 2021](#))

Helaas hebben wij geen

Violin chart ([Holtz, z.d.-j](#))

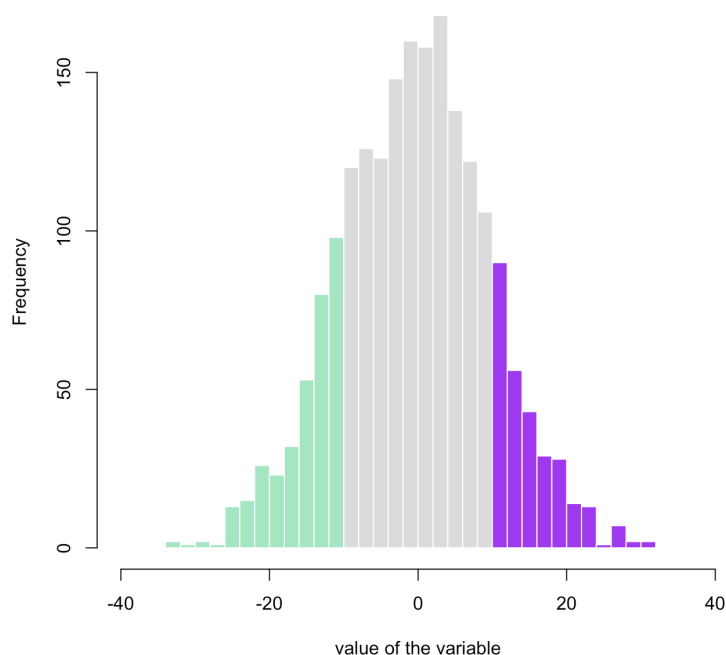


"Violin plots allow to visualize the distribution of a numeric variable for one or several groups. They are very well adapted for a large dataset."

Kortom, violin plots kunnen erg veel informatie weergeven, namelijk 3 verschillende waarden, wat erg handig is in ons project om bijvoorbeeld de gemaakte machines te weergeven mbt de tijd én alle machines in dit ene diagram te verwerken.

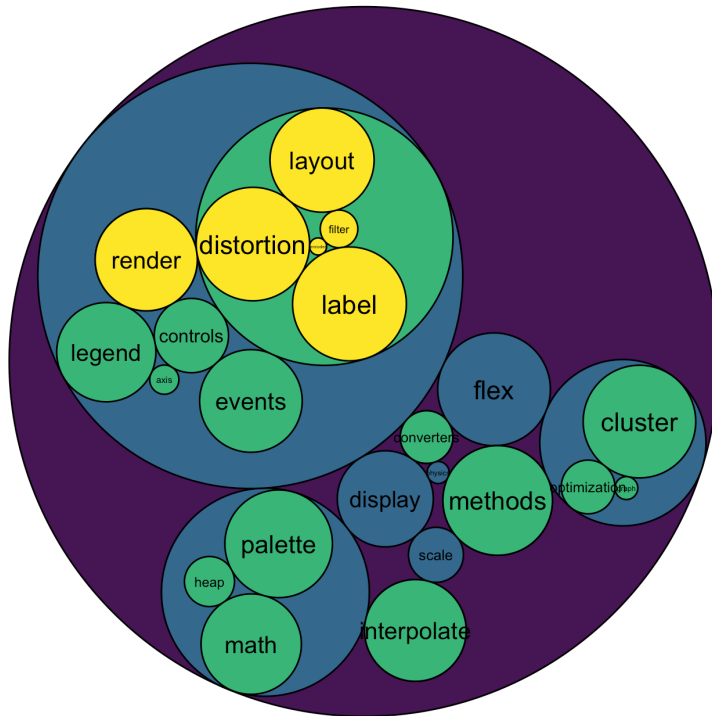
Het is niet een veelgebruikte grafiek en het vergt wat kennis en abstract denken om te lezen, maar als dit eenmaal duidelijk is leest het zo door. Vanwege het makkelijk verwerken en de andere hiervoor genoemde redenen hebben wij gekozen om ook de Violin chart verder uit te werken.

Histogram (Holtz, z.d.-f)



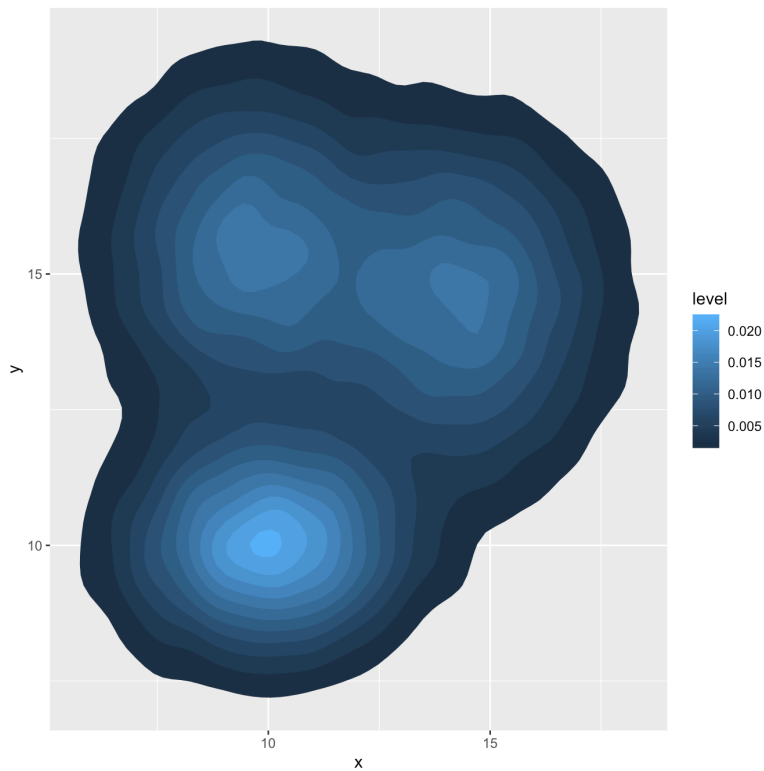
Een Histogram wordt meestal gemaakt als de verdeling ergens van weergegeven wordt en dit kan een erg handige grafiek zijn. Helaas is er geen functie gevonden die niet gedaan kan worden door de al gekozen grafieken, die daarnaast ook nog andere functies kunnen voldoen. Hierdoor is het histogram helaas afgefallen en wordt het niet verder uitgewerkt.

Circle packing (Holtz, z.d.-c)



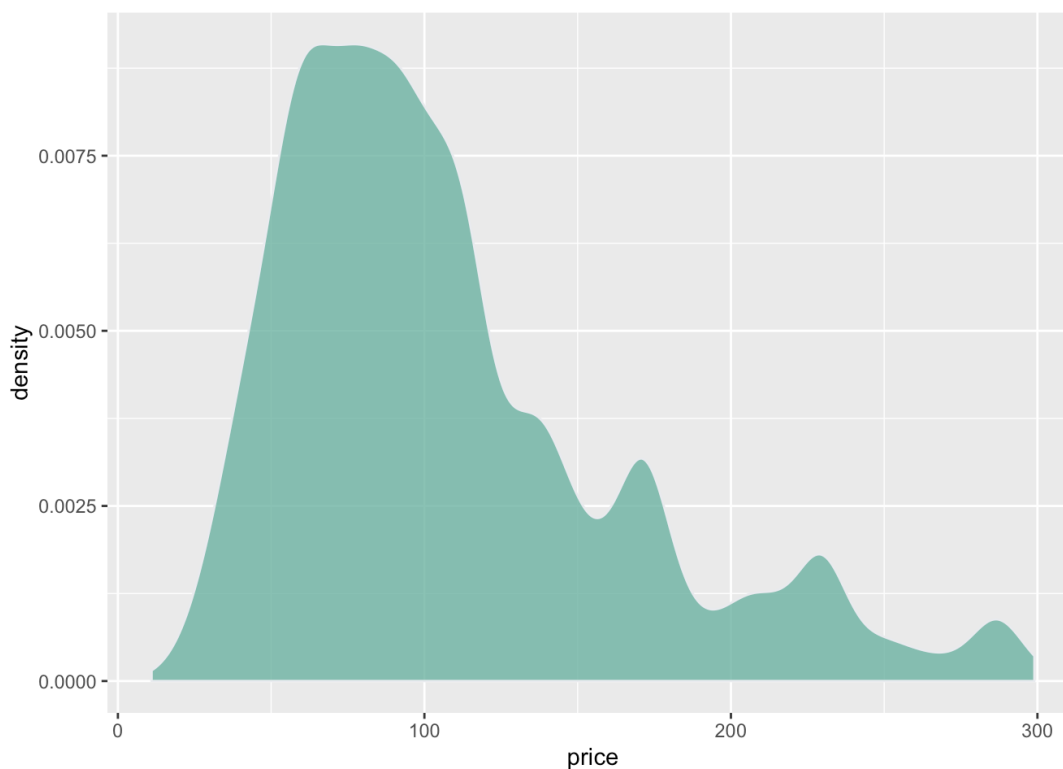
Een Circle packing is in hele specifieke gevallen best nuttig, maar helaas is het lastiger om te implementeren dan andere meer standaard grafieken. Daarnaast hebben we ook geen functie die een Circle packing beter kan visualiseren dan een andere grafiek, waardoor deze al vrij snel erbuiten valt.

Density 2d (Holtz, z.d.-a)



Net zoals bij de Circle Packing, hebben wij geen functie voor de Denisty 2d Graph binnen ons project en zal hier dus ook niet verder op ingegaan worden.

Density plot (Holtz, z.d.-d)



Als laatste hebben we de density plot. Wij hebben hierbij ook besloten dat we niet verder in gaan op de Density Plot. Dit komt omdat het hetzelfde kan en doet als een Line Graph, alleen hebben we democratisch besloten dat we de Line Graph mooier en beter vonden, dus werken we de Line Graph verder uit en niet de Density plot.

Hoofdvraag:

Welke grafieken gaan wij gebruiken voor ons rapport in R?

Bar graph (Holtz, z.d.-b)

- **Overzichtelijkheid**

Vanwege de hoge aanpasbaarheid van een Bar Graph (denk aan kleuren en nieuwe kolommen die makkelijk toegevoegd en verwijderd kunnen worden).

Daarnaast is het ook een grafiek die kleine veranderingen goed kan weergeven, omdat het erg makkelijk te vergelijken is.

- **Dekking van lading**

De Bar Graph is extreem makkelijk te lezen, omdat het simpele staven zijn zorgt het er niet voor dat er meer vragen opgeroepen worden bij de lezer.

Daarnaast wordt deze grafiek dus veel gebruikt in ons gekozen vakgebied, waar je de conclusie uit kan trekken dat deze grafiek toch echt nut heeft.

- **Misleidbaarheid**

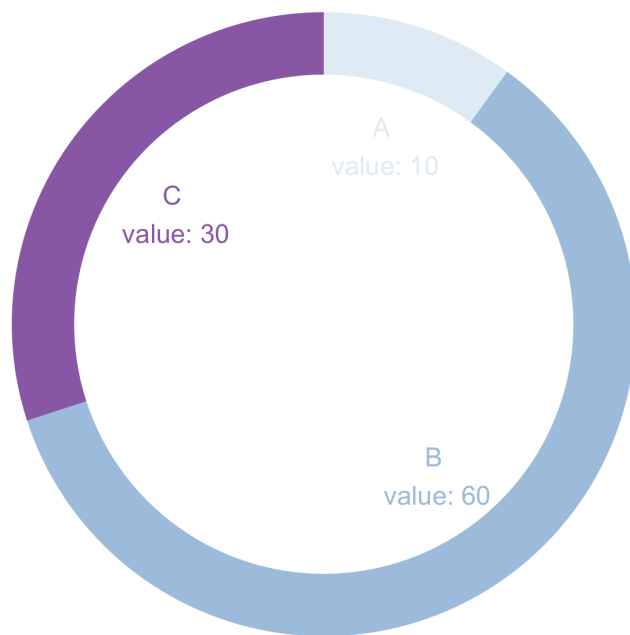
Vanwege het feit dat deze zo makkelijk te lezen is, is er eigenlijk geen mogelijkheid waarop dit diagram de lezer kan misleiden en zorgen dat er verkeerde conclusies uit getrokken wordt.

- **Gebruik binnen het project**

Aantal producten gemaakt x machineid (per 1). Hiermee kan je eenvoudig vergelijken welke machines erg veel moeten produceren en welke machines juist weinig produceren.



Donut chart (Holtz, z.d.-e)



- **Overzichtelijkheid**

Net zoals de Bar Graph is de Donut Graph erg overzichtelijk (mits hij minder dan 6 onderdelen bevat), zeker voor waar wij hem gaan gebruiken. Ons doel voor deze Graph is namelijk om hem te gebruiken om de run- en downtime te weergeven. De waardes hiervan worden ook weergegeven. Hiermee vermijd je het probleem dat toegelicht is in het kopje "Pie-chart" om te zorgen dat de grafiek optimaal gebruikt wordt.

- **Dekking van lading**

Zolang er gehouden wordt aan het bevatten van minder dan 6 onderdelen is het zo goed als gelijk duidelijk wat er mee bedoeld wordt en worden er geen vragen meer opgeroepen, waardoor dit een goede grafiek is om een specifieke lading te dekken.

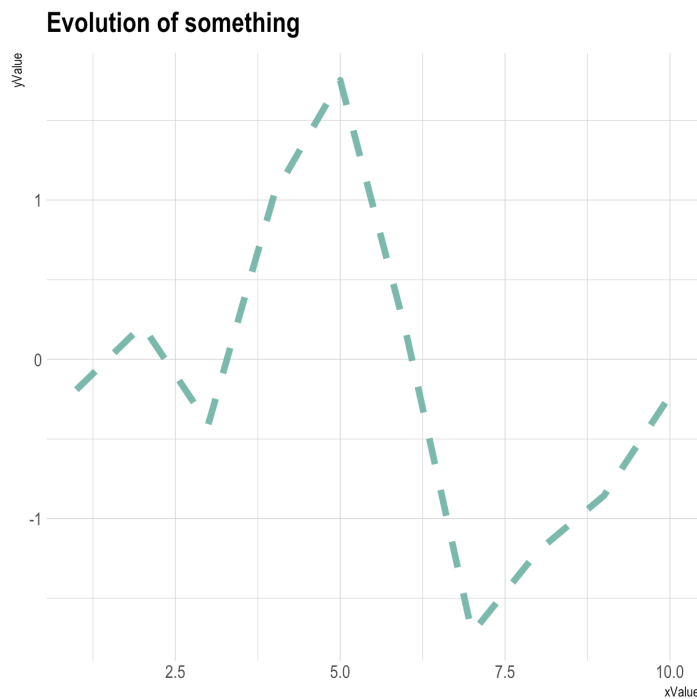
- **Misleidbaarheid**

Deze grafiek heeft een relatief grote kans om de lezer te misleiden als er geen waardes naastgezet zijn, omdat de verschillen dan alleen op het oog bepaald kunnen worden en dat is vrij moeilijk (zeker om accuraat te hebben) bij een cirkel.

- **Gebruik binnen het project**

Tijd dat een machine bezig is x tijd dat een machine uit staat. Als je zo'n grafiek voor alle machines maakt kan je erg makkelijk bepalen welke de zwakke schakel is, omdat deze zo goed als continu aan staat en dus het minste moet wachten op andere machines.

Line chart ([Holtz, z.d.-g](#))



- **Overzichtelijkheid**

De assen zijn erg gemakkelijk zelf in te stellen, waardoor de grafiek eigenlijk altijd een goede weergave geeft van de informatie en er niks verborgen blijft omdat het niet zichtbaar is.

- **Dekking van lading**

Het is een straight to the point grafiek die hierdoor de lading uitstekend dekt. Het wekt namelijk geen vragen meer op, vanwege de simpliciteit van deze grafiek.

- **Misleidbaarheid**

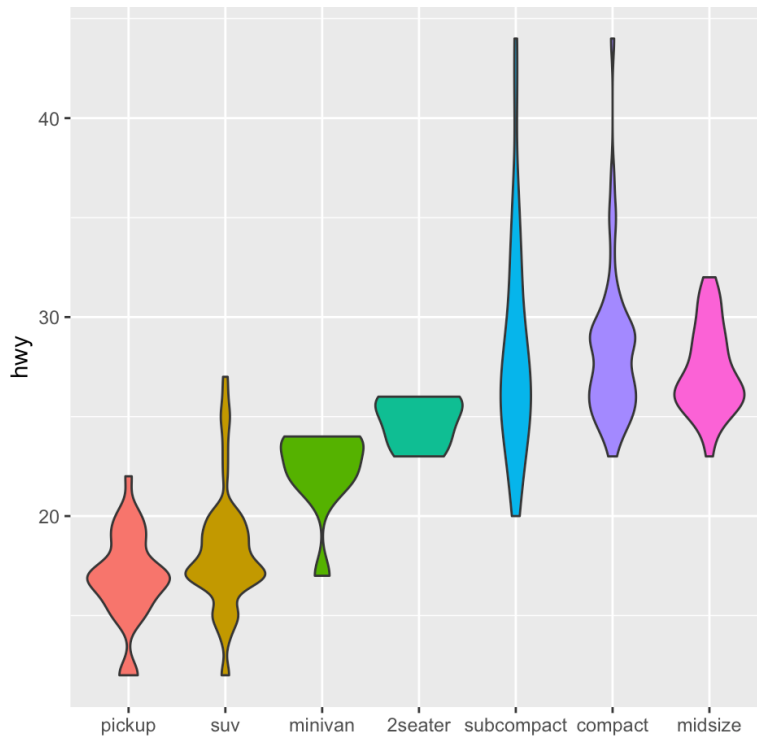
Het is zo ongeveer een standaard grafiek geworden die overal gebruikt kan worden, waardoor iedereen weet hoe je deze moet lezen. Ook is de informatie duidelijk en hoeven er geen aannames gemaakt te worden die voor onduidelijkheid of misleidingen zorgen.

- **Gebruik binnen het project**

Totaal geproduceerde producten x tijd om een beeld te krijgen van wanneer er hoeveel producten gemaakt worden, wat handig kan zijn om te zien of de productie ergens stopt.

Totaal geproduceerde producten x aantal dure machines i.p.v. goedkope (je begint met allemaal goedkope- en eindigt met allemaal dure machines). Hiermee kan je zien hoeveel effect het toevoegen van een dure machine heeft.

Violin chart (Holtz, z.d.-j)



- **Overzichtelijkheid**

Het is het minst overzichtelijke diagram van de 4 die wij hebben gekozen, maar als je er even naar kijkt wordt het al snel duidelijk wat het inhoudt en wat er wordt weergegeven met het diagram.

- **Dekking van lading**

Zoals eerder gezegd wordt het diagram een stuk duidelijker als het eenmaal klikt hoe het werkt. Daarna is het een geweldig diagram die meer lading bevat dan andere diagrammen, omdat er soort van 3 assen zijn, vanwege het feit dat de lijnen ook dikker of dunner kunnen worden. In het begin roept dit diagram aardig wat vragen op, maar minder dan bij bijvoorbeeld een density 2d graph, maar als je begrijpt hoe je het moet lezen wordt het ineens een redelijk eenvoudig diagram.

- **Misleidbaarheid**

Dit diagram heeft de mogelijkheid om het ergste de lezer te misleiden met informatie, omdat het wat onduidelijk kan zijn, maar als alle assen worden weergegeven wordt dit al geminimaliseerd.

Daarnaast is er veel informatie, maar het is wel goed gescheiden, waardoor het vrij moeilijk wordt om een verkeerde conclusie te trekken, mits je het diagram goed gelezen hebt.

- **Gebruik binnen het project**

Totaal geproduceerde producten x tijd x machine om een beeld te krijgen van wanneer elke machine hoeveel producten in de outputbuffer heeft, wat handig is om te zien wanneer in de tijdlijn een machine moet stoppen voor een andere.

Conclusie

Voor ons rapport gaan wij dus de volgende grafieken gebruiken: Bar graph, Donut chart, Line chart en Violin chart. Hiervoor gaan wij verschillende assen gebruiken per grafiek, namelijk:

- Bar graph

-Aantal producten gemaakt x machineld (per 1).



Voorbeeld voor *deluxe tafels*.

-(Aantal producten gemaakt / tijd laatste product gemaakt) x machineld.



Voorbeeld voor deluxe tafels.

- Donut chart
-Tijd dat een machine bezig is x tijd dat een machine niks aan het produceren is.



Voorbeeld voor deluxe tafels.

- Line chart
-Totaal geproduceerde producten x tijd.



Voorbeeld voor deluxe tafels.

- Violin chart
-Totaal geproduceerde producten x tijd x machine
(Bij nader inzien wordt deze grafiek geschrapt, omdat er een misverstand is in welke informatie hierin weergegeven werd. De informatie die wij hebben kan hier helaas dus niet in worden weergegeven, maar wordt als vervangen weergegeven in een linechart met Totaal geproduceerde producten x tijd.)

Met deze 5 verschillende grafiek templates gaan wij een rapport in R opstellen waar deze (sommige meerdere malen) in voorkomen.

Bibliografie

Holtz, Y. (z.d.-a). *2D Density Plot* | *The R Graph Gallery*. <https://r-graph-gallery.com/2d-density-chart.html>

Holtz, Y. (z.d.-b). *BarchaRT* | *The R Graph Gallery*. <https://r-graph-gallery.com/barplot.html>

Holtz, Y. (z.d.-c). *Customized circle packing with R and GGraph*. <https://r-graph-gallery.com/314-custom-circle-packing-with-several-levels.html>

Holtz, Y. (z.d.-d). *Density Chart* | *The R Graph Gallery*. <https://r-graph-gallery.com/density-plot.html>

Holtz, Y. (z.d.-e). *Donut Chart* | *The R Graph Gallery*. <https://r-graph-gallery.com/doughnut-plot.html>

Holtz, Y. (z.d.-f). *Histogram with colored tail*. <https://r-graph-gallery.com/83-histogram-with-colored-tail.html>

Holtz, Y. (z.d.-g). *Line Chart* | *The R Graph Gallery*. <https://r-graph-gallery.com/line-plot.html>

Holtz, Y. (z.d.-h). *Pie Chart* | *The R Graph Gallery*. <https://r-graph-gallery.com/pie-plot.html>

Holtz, Y. (z.d.-i). *The R Graph Gallery – help and inspiration for R charts*. The R Graph Gallery. <https://r-graph-gallery.com/index.html>

Holtz, Y. (z.d.-j). *TreeMap* | *The D3 Graph Gallery*. <https://r-graph-gallery.com/treemap.html>

Holtz, Y. (z.d.-k). *Violin Chart* | *The R Graph Gallery*. <https://r-graph-gallery.com/violin.html>

Oetting, J. (2023, 8 juni). 16 Best Types of Charts and Graphs for Data Visualization [+ Guide]. *HubSpot*. <https://blog.hubspot.com/marketing/types-of-graphs-for-data-visualization>

Rindhe, M. (2021, 14 december). Attractive data visualization: Get familiar with all popular graphs used in industry (Part-1). *Medium*. <https://medium.com/analytics-vidhya/attractive-data-visualization-get-familiar-with-all-popular-graphs-used-in-industry-part-1-b36fe0164e47>

Vogel, J. (z.d.-a). *ICT Research Methods — Methods Pack for research in ICT*. ICT Research Methods. <https://ictresearchmethods.nl/library/literature-study/>

Vogel, J. (z.d.-b). *ICT Research Methods — Methods Pack for research in ICT*. ICT Research Methods. <https://ictresearchmethods.nl/>

No labels

DDOA. Evaluate Confluence today.

This Confluence installation runs a Free Gliffy License - Evaluate the Gliffy
Confluence Plugin for your Wiki!