

Informatie Uitwisseling Visualisatie Opdracht

Matt ter Steege

m.j.ter.stege@students.uu.nl

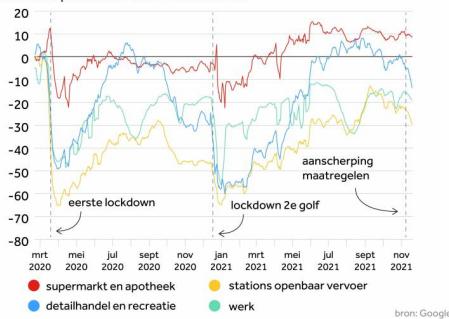
Universiteit Utrecht

Utrecht, Nederland

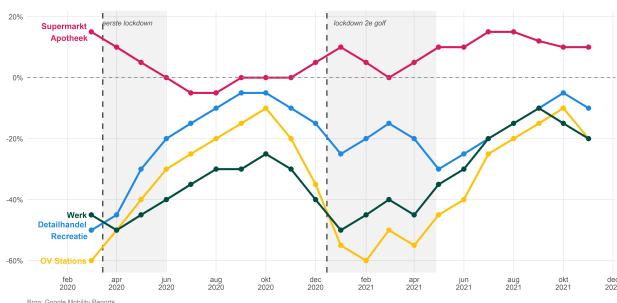
CONTENTS

Contents	1
1 Bezoekersgedrag Coronamaatregelen.....	2
2 Leesvaardigheidsscores Nederlandse scholieren.....	3
3 Zetelverdeling Tweede Kamer 2021 vs 2023	4
4 Autokleuren in Nederland.....	5
5 Gezondheidsrisico's bij twintigers	6
A Broncode grafiek 1.....	7
B Broncode grafiek 2.....	9
C Broncode grafiek 3.....	10
D Broncode grafiek 4.....	11
E Broncode grafiek 5	12

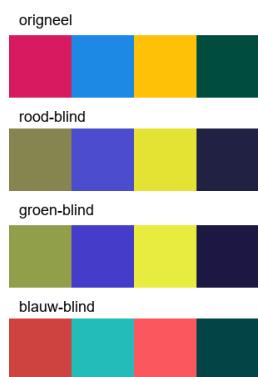
Deze golf blijven veel minder mensen thuis
Verschil ten opzichte van voor de coronacrisis



Verschil bezoekersaantal t.o.v. voor de coronacrisis



Figuur 1: Boven de visualisatie van de NOS, Onder de visualisatie van mij.



Figuur 2: Kleurenpalet voor verschillende kleurenziestoornissen

1 Bezoekersgedrag Coronamaatregelen (Foute kleurkeuze & te veel informatie)

Deze grafiek heb ik gevonden op de website van de NOS (Meindertsma, 2021). Ik wist dat er genoeg grafieken te vinden waren over het effect van de coronamaatregelen. Deze grafiek kwam redelijk snel naar en ik was er ook redelijk snel door overrompeld.

In de eerste visualisatie van de NOS gebeurt erg veel tegelijk. Als je naar een nieuwsbericht kijkt, dan wil je snel de belangrijkste informatie kunnen vinden. In deze grafiek is dat lastig, omdat er erg veel verschillende soorten informatie worden weergegeven. Het grootste probleem is dat in het algemeen de data veel te precies wordt weergegeven. Je ziet rustig 8 data punten per maand, terwijl dit niet nodig is om de trend te kunnen zien. Hierdoor wordt het erg druk en onoverzichtelijk. Ook is het kleurenschema niet heel goed gekozen. Er wordt gebruik gemaakt van rood, geel, blauw en groen, een **foutieve kleurcombinatie** voor mensen met kleurenblindheid. Dit is voor mensen met prot-, deuter- en tritanopie erg lastig te onderscheiden.

In mijn verbeterde grafiek heb ik gekozen voor een lijn grafiek, omdat ik dit voor de originele grafiek een erg goede keuze vond. Ik heb ook gespeeld met een heatmap, maar dit vond ik uiteindelijk toch minder overzichtelijk. Ik heb ervoor gekozen om de data uit te dunnen tot 1 datapunt per maand. Zoals Van der Rohe al zei: "Less is more". Hierdoor is de grafiek een stuk rustiger en overzichtelijker geworden. Daarnaast heb ik gekozen voor een kleurenschema dat goed bestand is tegen verschillende soorten kleurenblindheid. Ik heb het kleurenpalet op alle 3 de soorten getest en bij alle soorten bleven de kleuren goed te onderscheiden van elkaar.

Ik heb de lockdown periodes aangegeven met een lichte grijze balk op de achtergrond in plaats van een enkele lijn, dit geeft de daadwerkelijke periode beter weer. De "aanscherping maatregelen" lijn heb ik weggelegd, omdat de data daarna afgekapt werd en je dus geen conclusies kunt trekken over het effect hiervan op het bezoekersgedrag. De labels zijn naar de beginpunten van de lijnen verplaatst, zodat je niet steeds heen en weer hoeft te kijken tussen de lijnen en de legenda onderaan de grafiek.

Gemiddelde PISA-toetsscores leesvaardigheid

Periode 2003-2022

— OESO — EU14 — Nederland



Bron: Programme for International Student Assessment (PISA) • NRC 051223/Studio

2 Leesvaardigheidsscores Nederlandse scholieren (misleidende grafiek)

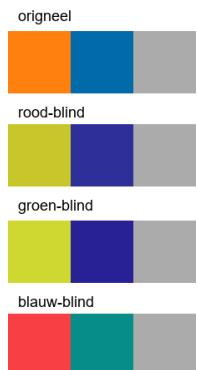
Deze bron heb ik gevonden op het NRC, een landelijke krant die hier verslag deed over de dalende leesvaardigheidsscores van de Nederlandse scholieren. (Kammer, 2023) Ik heb hem gevonden door te zoeken op onderwerpen die vaak grafisch worden weergegeven, zoals onderwijsresultaten.

Bij de linker (originele) grafiek is gekozen om aan te geven dat de dramatische keldering heeft gemaakt, echter is dit wel vertekend, omdat de grafiek niet op 0 begint, maar op 450. Dit is, als je er snel een blik op werpt zoals je in een krantenartikel vaak doet, erg **misleidend**. Niet alleen is de Y-as niet helemaal geweldig gekozen, ook op de X-as schaadt er wat. Hier is op het cruciale punt (vanaf het moment van schrijven) namelijk vorig jaar een gat van 4 in plaats van 3 jaar gemaakt. Dit wordt echter niet goed visueel gecommuniceerd, door bijvoorbeeld de X-as tussen die punten een jaar breder te maken. Zo kun je erg lastig redeneren of er tussen 2015 en 2018 dezelfde jaarlijkse daling is ingezet als tussen 2018 en 2022.

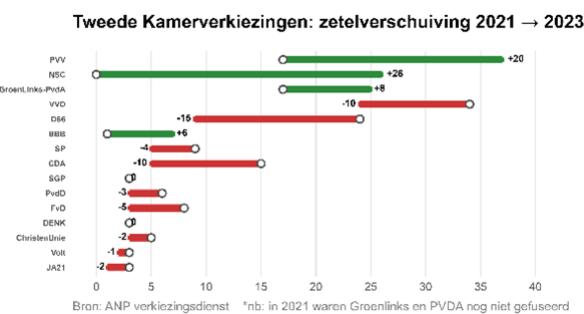
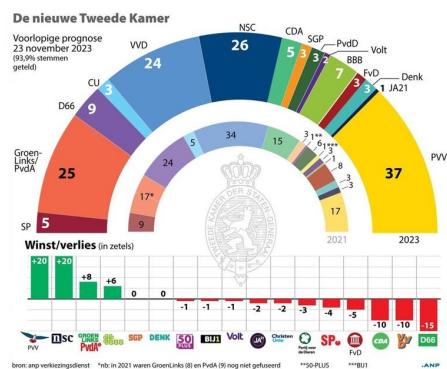
Als ik naar de grafiek kijk (en het bijbehorende artikel lees) dan is het vooral de bedoeling om de lezer te laten zien dat Nederland extreem slecht scoort in vergelijking met andere landen/regio's. Daarom heb ik in mijn verbeterde grafiek de focus gelegd op het vergelijken van de verschillende landen/regio's. Ik heb hier gekozen voor een barchart, omdat dit het makkelijkst is om verschillende landen met elkaar te vergelijken. De landen/regio's zijn op laag naar hoog in het eerste meetjaar (2003) gesorteerd, en deze volgorde is vastgehouden door de vervolgende jaren heen. Zo kun je makkelijk zien hoe landen/regio's zich ten opzichte van elkaar hebben ontwikkeld.

Ook heb ik hier gekozen voor een kleurenschema dat goed bestand is tegen verschillende soorten kleurenblindheid. Er is getest op Protan- Deuter- en Tritanopie en bij alle soorten bleven de kleuren goed te onderscheiden van elkaar. Daarnaast is de Y-as hier wel op 0 begonnen, zodat de verschillen in scores eerlijker worden weergegeven.

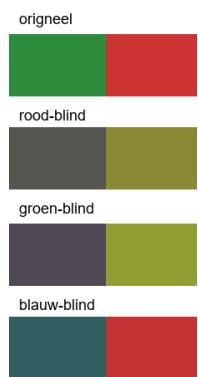
Figuur 3: Boven de visualisatie van het NRC, Onder de visualisatie van mij.



Figuur 4: Kleurenpalet voor verschillende kleurenzienstoornissen



Figuur 5: Boven de visualisatie van Oost, Onder de visualisatie van mij.



Figuur 6: Kleurenpalet voor verschillende kleurenziekenissen

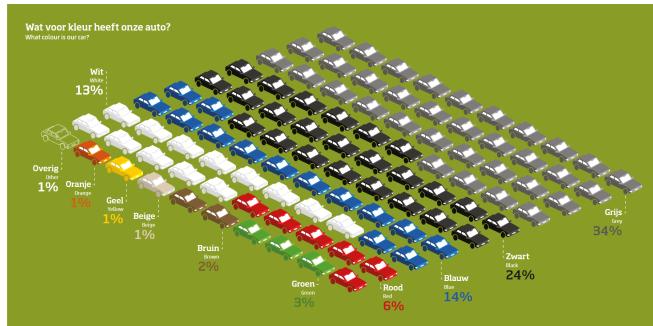
3 Zetelverdeling Tweede Kamer 2021 vs 2023 (Suboptimale channels & lage data-ink ratio)

De linker grafiek (tevens de originele grafiek) heb ik gevonden op de website van Oost, die verslag doet over voornamelijk Overijssel. (Oost, 2023)

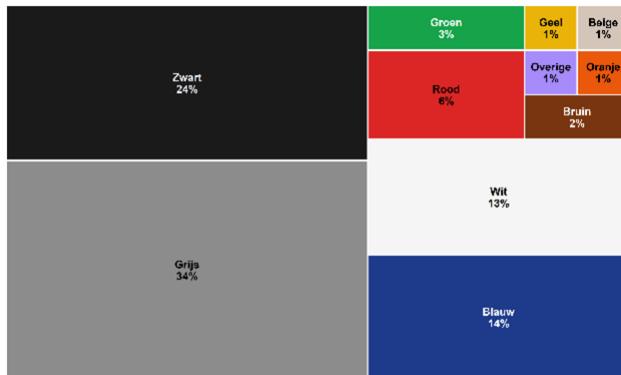
In de linker grafiek gebeurt heel veel. Je ziet een momentopname van 2021 en 2023 van de zetelverdeling van de Tweede Kamer. Maar ook de winst en verlies tussen de twee jaren. Deze grafiek gebruikt oppervlakte, ronding en ook verbergen ze de onzekerheid aangezien het om een prognose en 6% van de stemmen nog niet zijn meegenomen in de visualisatie. Deze grafiek gebruikt dus **suboptimale channels** om de data weer te geven. Oppervlakte is een slechte channel om hoeveelheden mee weer te geven, omdat mensen moeite hebben met het inschatten van oppervlakte. Daarnaast is ronding ook een slechte channel, omdat mensen moeite hebben met het inschatten van hoeken. Er staat weliswaar een label in of naast elke stuk van de halve cirkels, maar dit had al niet gehoeven als er betere channels waren gebruikt. Verder is de data-ink ratio toch tamelijk laag, omdat er erg veel poespas gebruikt is om hetzelfde aan te geven. Het wapen van Nederland is erg leuk, maar draagt niets bij aan het begrijpen van de data (of misschien wilde ze je gewoon afschrikken van de grafiek, wie weet).

Om deze grafiek te verbeteren heb ik in dit geval heb ik gekozen voor een barchart waarbij de start en eindposities de start en eindwaardes van de hoeveelheid zetels die wordt (verwacht) behaald te worden. Ik heb hier gekozen voor een dubbele encoding. Als eerste de kleur van de balken, in dit geval het rood/groen kleurenschema. Deze is wel slechter te zien door mensen met Protanomalias of Protanopie, maar aangezien er tegelijkertijd gebruik wordt gemaakt van de richting van de balk om aan te geven of er een positieve of negatieve verschuiving in het behaalde aantal zetels is, kan deze grafiek ook makkelijk door mensen met (een vorm van) kleurenblindheid bekijken worden.

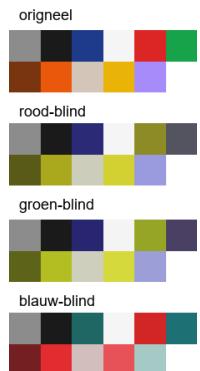
Ik heb alleen helaas niet kunnen vinden wat de onzekerheid van de grafiek was, aangezien deze nu ook in mijn verbeterde grafiek niet aanwezig is. Had ik deze wel gevonden, dan was deze in schuifgestreepte manier terug te vinden aan het einde van de balk in dezelfde kleur als de balk zelf.



Autokleuren in Nederland



Figuur 7: Boven de visualisatie van het CBS, Onder de visualisatie van mij.



Figuur 8: Kleurenpalet voor verschillende kleurenzienstoornissen

4 Autokleuren in Nederland (lage data-ink ratio)

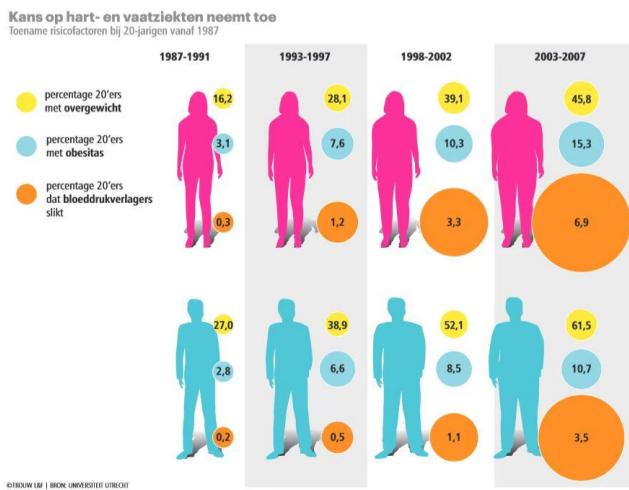
De linker grafiek van het CBS (CBS, 2022) laat de verdeling van autokleuren in Nederland zien. Er gebeurt te veel tegelijk. De data wordt weergegeven met kleine autootjes die in schuine rijen liggen, met een duidelijk perspectief en een soort 3D-effect. Er is duidelijk een visualisatietijger een goede woensdagmiddag mee bezig geweest en het ziet er leuk uit. Hierdoor wordt het helaas wel lastiger dan nodig om de data af te lezen. Hoeveel scheelt blauw nu precies met wit?

Daarnaast is de **data-ink ratio laag**. Een groot deel van de gebruikte inkt gaat op aan decoratie: de autovormmpjes, de hellende compositie, de achtergrondkleur en de schaduwverwerking. Al die elementen dragen niets bij aan het beantwoorden van de vraag welke autokleuren dominant zijn in Nederland. Ze leiden af en maken het lastiger om patronen te herkennen. De grafische integriteit komt daarmee onder druk te staan, omdat de visuele impact niet meer in verhouding staat tot de feitelijke data.

Om dit te verbeteren heb ik gekozen voor een veel eenvoudigere visualisatie, namelijk een vlakke, overzichtelijke verdeling waarin elke kleur een duidelijk afgebakend vlak krijgt met het bijbehorende percentage. In deze grafiek wordt alleen oppervlakte gebruikt, zonder perspectief of diepte. De verhoudingen zijn direct zichtbaar en vergelijkbaar. Grijs is onmiskenbaar het grootst, gevolgd door zwart, en daarna pas blauw en wit. De kleinere categorieën verdwijnen niet, maar eisen ook niet onnodig aandacht op.

Door alle niet-essentiële elementen weg te laten, is de data-ink ratio aanzienlijk hoger. Vrijwel alles wat je ziet, is data. Er is geen poespas, geen decoratie en geen visuele ruis. Je krijgt zo in één oogopslag inzicht in de verdeling van autokleuren in Nederland. Dat maakt deze simpele grafiek niet alleen rustiger, maar vooral eerlijker en effectiever dan de originele.

Deze grafiek is helaas niet heel goed te bekijken door mensen met een vorm van kleurenblindheid. Echter is hiertegen wel te argumenteren dat de mensen met kleurenblindheid waarschijnlijk ook niet heel erg geïnteresseerd zijn in autokleuren, aangezien zij deze zelf ook niet goed kunnen onderscheiden. Ook zou het niet logisch zijn om te proberen een kleurenpalet te maken dat voor alle vormen van kleurenblindheid goed te onderscheiden is, want dan wordt het lastiger voor mensen zonder kleurenblindheid om de kleuren te linken aan de daadwerkelijke autokleuren.



Figuur 9: Boven de visualisatie van Trouw, Onder de visualisatie van mij.

5 Gezondheidsrisico's bij twintigers (misleidende grafiek)

(Kuiken, 2016) heeft deze kleurrijke grafiek in de Trouw gepubliceerd. Hij laat de toename van gezondheidsrisico's onder twintigers zien, uitgesplitst naar geslacht.

In de originele grafiek werd gewerkt met twee rijen poppetjes, blauw voor mannen en roze voor vrouwen. Elk poppetje had drie bolletjes die de drie gezondheidsrisico's moesten voorstellen. Op het eerste gezicht oogt dat speels, maar zodra je probeert te begrijpen hoeveel iets toeneemt, dan wordt het lastig. De bolletjes vertegenwoordigen percentages, maar die percentages zijn niet visueel in verhouding. De cirkels zijn met bijvoorbeeld 45.8% misschien (want cirkels aflezen is lastig...) 10x kleiner dan 6.9 procent. Dat is natuurlijk een extreem misleidend beeld.

Daarnaast zijn de cirkels die wél met elkaar in verhouding zijn ook erg lastig met elkaar te vergelijken, omdat de gele cirkels bij de mannen tussen de twee einde verdubbeld, terwijl ze erg bijna hetzelfde uitzien. Dit is voor Trouw natuurlijk een mooie manier om het dramatische karakter van de toename te benadrukken, maar het schaadt de grafische integriteit. Voor lezers is het een aardig **misleidende** grafiek.

Daarnaast is de data-ink ratio laag: veel inkt gaat naar poppetjes, kleuren en decoratie, terwijl de daadwerkelijke informatie amper preciezer wordt.

In de verbeterde grafiek heb ik dit volledig omgedraaid. Ik heb gekozen voor lijndiagrammen, uitgesplitst per gezondheidsrisico: bloeddrukverlagers, obesitas en overgewicht. De tijd staat op de horizontale as, het percentage op de verticale as. Mannen en vrouwen worden onderscheiden met kleur, maar delen exact dezelfde schaal. Daardoor ontstaat meteen duidelijkheid. Ik zie niet alleen dat elk risico toeneemt, maar ook hoe snel, hoe sterk en wie er hoger zit.

De belangrijkste channel wordt ingezet: positie op een gemeenschappelijke schaal. De lijnen maken trends zichtbaar zonder extra uitleg. Vooral bij overgewicht springt het verschil tussen mannen en vrouwen er direct uit. Dat hoeft je niet te interpreteren, dat zie je gewoon.



Figuur 10: Kleurenpalet voor verschillende kleurenzienvoorrnissen

References

- CBS. (2022). *Wat voor kleur heeft onze auto?* CBS. <https://longreads.cbs.nl/nederland-in-cijfers-2022/wat-voor-kleur-heeft-onze-auto/>
- Kammer, C. (2023, December). *Bij lezen staat nederland nu bijna onderaan in de eu.* NRC. <https://www.nrc.nl/nieuws/2023/12/05/bij-lezen-staat-nederland-in-eu-alleen-nog-boven-griekenland-a4183300>
- Kuiken, A. (2016, April). Trouw. <https://www.trouw.nl/nieuws/iedere-generatie-wordt-weer-een-beetje-dikker-en-ongezonder-bd1c3910/>
- Meindertsma, B. (2021, November). *Deze grafieken helpen je de coronagolf beter te begrijpen.* NOS. <https://nos.nl/artikel/2407066-deze-grafieken-helpen-je-de-coronagolf-beter-te-begrijpen>
- Oost. (2023, November). *Er is een winnaar, nu nog een kabinet: Dit gebeurt er de komende weken.* Oost. <https://www.oost.nl/nieuws/2286898/er-is-een-winnaar-nu-nog-een-kabinet-dit-gebeurt-er-de-komende-weken>

A Broncode grafiek 1

```

1 library(ggplot2)
2 library(tidyr)
3 library(dplyr)
4 library(readr)
5
6 # Data inline
7 data <- data.frame(
8   Maand = c("mrt 2020", "apr 2020", "mei 2020", "jun 2020", "jul 2020", "aug 2020",
9     "sep 2020", "okt 2020", "nov 2020", "dec 2020", "jan 2021", "feb 2021",
10    "mrt 2021", "apr 2021", "mei 2021", "jun 2021", "jul 2021", "aug 2021",
11    "sep 2021", "okt 2021", "nov 2021"),
12   `Supermarkt en apotheek` = c(15, 10, 5, 0, -5, -5, 0, 0, 0, 5, 10, 5, 0, 5, 10, 10, 15, 15, 12, 10, 10),
13   `Detailhandel en recreatie` = c(-50, -45, -30, -20, -15, -10, -5, -5, -10, -15, -25, -20, -15, -20, -30, -25, -20, -15,
14     -10, -5, -10),
15   `Stations openbaar vervoer` = c(-60, -50, -40, -30, -25, -20, -15, -10, -20, -35, -55, -60, -50, -55, -45, -40, -25,
16     -20, -15, -10, -20),
17   Werk = c(-45, -50, -45, -40, -35, -30, -25, -30, -40, -50, -45, -40, -45, -35, -30, -20, -15, -10, -15, -20),
18   check.names = FALSE
19 )
20
21 # Verwijder deze regel:
22 # data = pd.read_csv('mobility_data.csv')
23
24 # Data omzetten naar long format
25 data_long <- data %>%
26   pivot_longer(
27     cols = -Maand,
28     names_to = "Categorie",
29     values_to = "Percentage"
30   ) %>%
31   mutate(
32     Maand_date = as.Date(paste0(Maand, " 01")), format = "%b %Y %d"),
33     Categorie_kort = case_when(
34       Categorie == "Supermarkt en apotheek" ~ "Supermarkt\nApotheek",
35       Categorie == "Detailhandel en recreatie" ~ "Detailhandel\nRecreatie",
36       Categorie == "Stations openbaar vervoer" ~ "OV Stations",
37       Categorie == "Werk" ~ "Werk "
38     )
39
40 # Kleuren zoals in de originele grafiek
41 kleuren <- c(
42   "Supermarkt\nApotheek" = "#D81B60",
43   "Detailhandel\nRecreatie" = "#1E88E5",
44   "OV Stations" = "#FFC107",
45   "Werk" = "#004D40"
46 )
47
48 # Lockdown periodes defini ren
49 lockdown1_start <- as.Date("2020-03-15")
50 lockdown1_end <- as.Date("2020-05-31")
51 lockdown2_start <- as.Date("2020-12-15")
52 lockdown2_end <- as.Date("2021-04-28")
53
54 p2 <- ggplot(data_long, aes(x = Maand_date, y = Percentage, color = Categorie_kort, group = Categorie_kort)) +
55   # Lockdown achtergronden
56   annotate("rect", xmin = lockdown1_start, xmax = lockdown1_end,
57             ymin = -Inf, ymax = Inf, alpha = 0.1, fill = "gray50") +
58   annotate("rect", xmin = lockdown2_start, xmax = lockdown2_end,
59             ymin = -Inf, ymax = Inf, alpha = 0.1, fill = "gray50") +
60   # Verticale lijnen voor lockdown starts
61   geom_vline(xintercept = as.numeric(lockdown1_start), linetype = "dashed",
62             color = "gray30", linewidth = 0.8) +
63   geom_vline(xintercept = as.numeric(lockdown2_start), linetype = "dashed",
64             color = "gray30", linewidth = 0.8) +
65   # Data lijnen
66   geom_line(linewidth = 1.2) +
67   geom_point(size = 2.5) +
68   geom_hline(yintercept = 0, linetype = "dashed", color = "grey50", linewidth = 0.5) +
69   # Labels aan het begin van de lijnen met percentage
70   geom_text(data = data_long %>%

```

```
70     group_by(Categorie_kort) %>%
71       slice_min(Maand_date, n = 1),
72       aes(label = paste0(Categorie_kort)),
73       hjust = 1.1, vjust = 0.5, size = 3.5, fontface = "bold") +
74 # Lockdown labels
75   annotate("text", x = lockdown1_start + 30, y = 18,
76             label = "eerste lockdown", size = 3, color = "gray30", fontface = "italic") +
77   annotate("text", x = lockdown2_start + 40, y = 18,
78             label = "lockdown 2e golf", size = 3, color = "gray30", fontface = "italic") +
79   scale_color_manual(values = kleuren) +
80   scale_x_date(date_breaks = "2 months", date_labels = "%b\n%Y",
81               limits = c(as.Date("2020-01-15"), max(data_long$Maand_date))) +
82   scale_y_continuous(labels = function(x) paste0(x, "%")) +
83   labs(
84     title = "Verschil bezoekersaantal t.o.v. voor de coronacrisis",
85     subtitle = "",
86     caption = "Bron: Google Mobility Reports",
87     x = NULL,
88     y = NULL
89   ) +
90   theme_minimal(base_size = 11) +
91   theme(
92     plot.title = element_text(face = "bold", size = 14, margin = margin(b = 5)),
93     plot.subtitle = element_text(size = 11, color = "grey40", margin = margin(b = 15)),
94     plot.caption = element_text(size = 8, color = "grey50", hjust = 0, margin = margin(t = 10)),
95     plot.margin = margin(15, 35, 15, 15),
96     panel.grid.minor = element_blank(),
97     legend.position = "none"
98   )
99
100 ggsave("plot2_alle_lijen.png", p2, width = 11, height = 6, dpi = 600, bg = "white")
```

Listing 1: R code voor bezoekersgedrag coronamaatregelen

B Broncode grafiek 2

```

1 library(tidyverse)
2
3 # Data invoeren
4 data <- tibble(
5   Jaar = c(2003, 2006, 2009, 2012, 2015, 2018, 2022),
6   OESO = c(494, 485, 490, 493, 490, 487, 476),
7   EU14 = c(498, 493, 495, 499, 499, 493, 480),
8   Nederland = c(513, 507, 508, 511, 503, 485, 459)
9 )
10
11 # Data transformeren naar long format
12 data_long <- data %>%
13   pivot_longer(
14     cols = c(OESO, EU14, Nederland),
15     names_to = "Regio",
16     values_to = "Waarde"
17   ) %>%
18   mutate(
19     Regio = factor(Regio, levels = c("OESO", "EU14", "Nederland")),
20     Jaar = factor(Jaar)
21   )
22
23 # Kleuren consistent met zachteblauwe tint
24 kleuren <- c("OESO" = "#ff800f", "EU14" = "#016aa8", "Nederland" = "#ababab")
25
26 ggplot(data_long, aes(x = Jaar, y = Waarde, fill = Regio)) +
27   geom_col(position = position_dodge(width = 0.8), width = 0.7) +
28   scale_fill_manual(values = kleuren) +
29   scale_y_continuous(
30     limits = c(0, 520),
31     breaks = seq(0, 520, 100),
32     expand = expansion(mult = c(0, 0.02))
33   ) +
34   labs(
35     title = "Vergelijking scores over tijd",
36     subtitle = "Nederland scoort hoger dan EU14 en OESO tot 2018",
37     x = "Jaar",
38     y = "Score",
39     fill = NULL,
40     caption = "Grouped bar chart: maakt directe vergelijking tussen groepen per jaar makkelijk"
41   ) +
42   theme_minimal(base_size = 11) +
43   theme(
44     plot.title = element_text(face = "bold", size = 13, margin = margin(b = 5)),
45     plot.subtitle = element_text(size = 10, color = "grey40", margin = margin(b = 15)),
46     plot.caption = element_text(size = 8, color = "grey50", hjust = 0, margin = margin(t = 10)),
47     axis.text = element_text(size = 9, color = "grey40"),
48     axis.title = element_text(size = 10, color = "grey20"),
49     axis.title.x = element_text(margin = margin(t = 10)),
50     axis.title.y = element_text(margin = margin(r = 10)),
51     panel.grid.minor = element_blank(),
52     panel.grid.major.x = element_blank(),
53     panel.grid.major.y = element_line(color = "grey90", linewidth = 0.3),
54     legend.position = "top",
55     legend.justification = "left",
56     legend.text = element_text(size = 10),
57     plot.margin = margin(15, 15, 15, 15)
58   )
59
60 ggsave("grafiek.png", width = 8, height = 5, dpi = 600, bg = "white")

```

Listing 2: R code voor leesvaardigheidsscores Nederlandse scholieren

C Broncode grafiek 3

De broncode voor grafiek 3 is helaas verloren gegaan.

D Broncode grafiek 4

```

1 library(ggplot2)
2 library(dplyr)
3 library(forcats)
4 library(ggrepel)
5
6 # Data invoeren
7 data <- data.frame(
8   Kleur = c("Overige", "Oranje", "Geel", "Beige", "Bruin", "Groen",
9           "Rood", "Wit", "Blauw", "Zwart", "Grijs"),
10  Percentage = c(1, 1, 1, 1, 2, 3, 6, 13, 14, 24, 34)
11 )
12
13 # Kleuren defini ren (zo realistisch mogelijk)
14 kleur_mapping <- c(
15   "Grijs" = "#8C8C8C",
16   "Zwart" = "#1a1a1a",
17   "Blauw" = "#1E3A8A",
18   "Wit" = "#F5F5F5",
19   "Rood" = "#DC2626",
20   "Groen" = "#16A34A",
21   "Bruin" = "#78350F",
22   "Beige" = "#D4C5B9",
23   "Geel" = "#EAB308",
24   "Oranje" = "#EA580C",
25   "Overige" = "#A78BFA"
26 )
27
28 # Sorteer data
29 data <- data %>%
30   mutate(Kleur = fct_reorder(Kleur, Percentage))
31
32 # Voor treemap moeten we wat berekenen
33 library(treemapify)
34
35 data_tree <- data %>%
36   mutate(
37     label = paste0(Kleur, "\n", Percentage, "%"),
38     text_color = ifelse(Kleur %in% c("Zwart", "Blauw", "Bruin", "Groen"), "white", "black")
39   )
40
41 p2 <- ggplot(data_tree, aes(area = Percentage, fill = Kleur, label = label)) +
42   geom_treemap(color = "white", size = 2) +
43   geom_treemap_text(
44     aes(color = text_color),
45     place = "centre",
46     size = 12,
47     fontface = "bold",
48     lineheight = 0.9
49   ) +
50   scale_fill_manual(values = kleur_mapping) +
51   scale_color_identity() +
52   labs(
53     title = "Autokleuren in Nederland",
54     subtitle = "",
55     caption = ""
56   ) +
57   theme_minimal(base_size = 11) +
58   theme(
59     plot.title = element_text(face = "bold", size = 14, margin = margin(b = 5)),
60     plot.subtitle = element_text(size = 11, color = "grey40", margin = margin(b = 15)),
61     plot.caption = element_text(size = 8, color = "grey50", hjust = 0, margin = margin(t = 10)),
62     plot.margin = margin(15, 15, 15, 15),
63     legend.position = "none",
64     axis.text = element_blank(),
65     axis.title = element_blank(),
66     panel.grid = element_blank()
67   )
68
69 ggsave("autokleur_alt2_treemap.png", p2, width = 10, height = 7, dpi = 600, bg = "white")

```

Listing 3: R code voor autokleuren in Nederland

E Broncode grafiek 5

```

1 library(ggplot2)
2 library(tidyr)
3 library(dplyr)
4
5 data <- data.frame(
6   Periode = c("1987-1991", "1993-1997", "1998-2002", "2003-2007"),
7   Vrouwen_Overgewicht = c(16.2, 28.1, 39.1, 45.8),
8   Vrouwen_Obesitas = c(3.1, 7.6, 10.3, 15.3),
9   Vrouwen_Bloeddrukverlagers = c(0.3, 1.2, 3.3, 6.9),
10  Mannen_Overgewicht = c(27, 38.9, 52.1, 61.5),
11  Mannen_Obesitas = c(2.8, 6.6, 8.5, 10.7),
12  Mannen_Bloeddrukverlagers = c(0.2, 0.5, 1.1, 3.5)
13 )
14
15 # Data omzetten naar long format
16 data_long <- data %>%
17   pivot_longer(
18     cols = -Periode,
19     names_to = c("Geslacht", "Categorie"),
20     names_sep = "_",
21     values_to = "Percentage"
22   ) %>%
23   mutate(
24     Periode_num = case_when(
25       Periode == "1987-1991" ~ 1989,
26       Periode == "1993-1997" ~ 1995,
27       Periode == "1998-2002" ~ 2000,
28       Periode == "2003-2007" ~ 2005
29     )
30   )
31
32 # Kleuren
33 kleuren <- c(
34   "Overgewicht" = "#E69F00",
35   "Obesitas" = "#D55E00",
36   "Bloeddrukverlagers" = "#0072B2"
37 )
38
39 kleuren_geslacht <- c("Vrouwen" = "#CC79A7", "Mannen" = "#009E73")
40
41 p4 <- ggplot(data_long, aes(x = Periode_num, y = Percentage, color = Geslacht, group = Geslacht)) +
42   geom_line(linewidth = 1.2) +
43   geom_point(size = 3) +
44   facet_wrap(~Categorie, ncol = 3) +
45   scale_color_manual(values = kleuren_geslacht) +
46   scale_x_continuous(
47     breaks = c(1989, 1995, 2000, 2005),
48     labels = c("'87-'91", "'93-'97", "'98-'02", "'03-'07")
49   ) +
50   scale_y_continuous(labels = function(x) paste0(x, "%")) +
51   labs(
52     title = "Toenemende gezondheidsrisico's 20ers",
53     subtitle = "",
54     caption = "",
55     x = NULL,
56     y = "",
57     color = "Geslacht"
58   ) +
59   theme_minimal(base_size = 11) +
60   theme(
61     plot.title = element_text(face = "bold", size = 14, margin = margin(b = 5)),
62     plot.subtitle = element_text(size = 11, color = "grey40", margin = margin(b = 15)),
63     plot.caption = element_text(size = 8, color = "grey50", hjust = 0, margin = margin(t = 10)),
64     plot.margin = margin(15, 15, 15, 15),
65     panel.grid.minor = element_blank(),
66     strip.text = element_text(face = "bold", size = 11),
67     legend.position = "bottom"
68   )
69
70 ggsave("plot4_small_multiples.png", p4, width = 11, height = 5, dpi = 600, bg = "white")

```

Listing 4: R code voor gezondheidsrisico's bij twintigers