



WAT IS DATADOCUMENTATIE?



DATADOCUMENTATIE

OPSLAAN, ARCHIVEREN & PUBLICEREN

PROJECT

DATA

TOEGANG & DELEN



OPSLAAN, ARCHIVEREN EN PUBLICEREN

De vindbaarheid, toegankelijkheid en herbruikbaarheid van de data wordt door het goed documenteren verhoogd, zowel bij de opslag tijdens het project als bij het archiveren en publiceren.

- stel de [mapstructuur](#) en [bestandsnamen](#) vast zodat de data in het onderzoeksproject efficiënt te vinden en herkennen zijn
- bepaal waar je efficiënt je datadocumentatie tijdens je onderzoek vastlegt en verzamelt
- hou de datadocumentatie bij tijdens je onderzoek. Zorg dat je een procedure maakt, zodat je op de juiste momenten hier aan herinnerd wordt en aan werkt
- werk tijdens je onderzoek je [datamanagementplan](#) steeds bij





PROJECT

Informatie over het project.

- [Projectgegevens](#)
 - medewerkers
 - instelling
- [Financier](#)
- [Gerelateerde projecten en publicaties](#)
 - projecten
 - publicaties uit de data
- [Datamanagementplan](#)





DATA

Beschrijving van ontstaan en bewerking van dataset.

- Omschrijving variabelen
 - [Codeboek](#)
 - [Scripts](#)
 - [SPSS en ATLAS.ti](#)
- Omschrijving analyse
 - [SPSS \(Syntax\)](#)
 - [R](#)
 - [Python](#)
- [Software](#) ontwikkeld voor dit onderzoek
- [Onderzoeksprotocol](#)
- [Informed consent](#)
- [Informatiebrief](#)
- [Vragenlijst](#)
- [Stimuli](#)





TOEGANG EN DELEN

Informatie over hoe de onderzoeksdata toegankelijk wordt gemaakt.

- [README-file](#) over structuur bestanden
- [Metadatastandaard](#)
- [Vocabulaire](#)
- [Persistent Identifier](#)
- [Licenties](#)
- [Bestandsformaten](#)





PROJECTGEGEVENS

- naam project Nederlandstalig en Engelstalig
- afkorting/acroniem project
- korte samenvatting project
- looptijd project
- naam en contactgegevens hoofdonderzoeker/ORCID en/of DAI
- overige projectleden/rol/ORCID en/of DAI
- instelling
- lectoraat
- discipline
- trefwoorden project Nederlandstalig en Engelstalig





- naam financier
- naam subsidie
- project/subsidienummer





GERELATEERDE PROJECTEN – PUBLICATIES

Hier noem je publicaties of projecten die de onderzoeksdata gebruiken.

- vermeld de volledige referentie van de publicatie

Voorbeeld:

Bovens, J., Moresi, S., Van den Langenberg, N., & Snoeren, M. (2022). De sleutel naar grensoverstijgend samenwerken in learning communities: De rol van bruggenbouwer. *Tijdschrift voor HRM*, 25(3), 85-103. <https://doi.org/10.5117/THRM2022.3.003.BOVE>





DATAMANAGEMENTPLAN

Efficiënt omgaan met onderzoeksgegevens vereist planning. Steeds meer subsidieverstrekkers stellen eisen aan datamanagement en ook de Nederlandse gedragscode wetenschappelijke integriteit hanteert richtlijnen voor goed databeheer. Het helpt onderzoekers echter ook bij het inventariseren van risico's bij het beheer van onderzoeksgegevens gedurende het hele onderzoeksproces. Het datamanagementplan (DMP) is een levend document dat je voortdurend blijft updaten. Pas na afloop van je onderzoeksproject is het definitief. Een DMP schrijf je altijd met ondersteuning van een datasteward. De volgende punten komen aan bod in een DMP.

- administratieve gegevens
- beschrijving van de dataverzameling
- standaarden en metadata
- ethisch en juridisch
- opslag en autorisatie
- archivering en hergebruik
- bewaartermijn
- beschrijving software en tools



OMSCHRIJVING VARIABELEN CODEBOEK

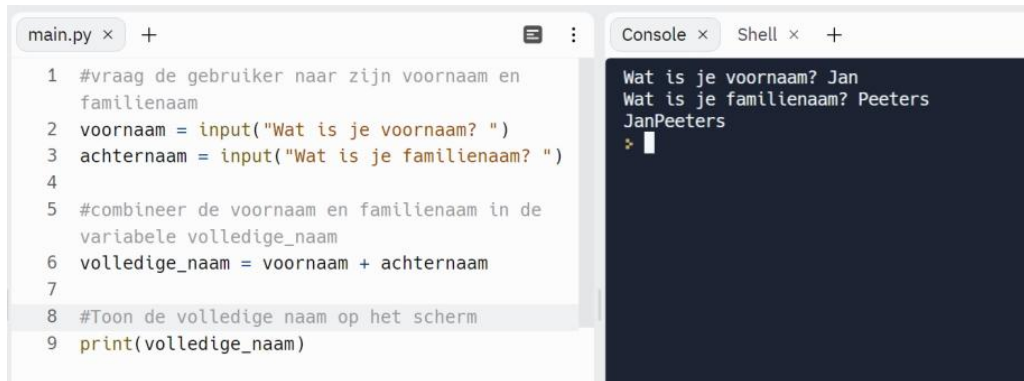
- Contextbeschrijving: wanneer, waar en door wie
- Beschrijving van onderzoekspopulatie
- Selectiemethode
- Tool voor dataverzameling
- Codering van de variabelen
 - Naam en labels van de variabelen van de dataset
 - Codering antwoordopties
- Gerelateerde publicaties
- Meeteenheden
- Uitleg van afkortingen, concepten, categorieën
- Uitleg van codes en symbolen die gebruikt worden bij missing data
- Frequentietellingen, aantekeningen
- Voorbeelden: readable codebook in PDF format. Dit is hetzelfde codeboek in machine-actionable DDI-formaat (XML-formaat). Deze bestanden zijn gegenereerd met behulp van de DDI-Codebook-standaard en zijn gemarkeerd met Nesstar Publisher. Meer voorbeelden [hier](#)

Leefbaar- heid dorp Q	Variabele omschrijving	VAR	Score Likert- schaal	Scores				
Kolom 1	tevredenheid tijdigheid OV	VAR 1	1 t/m 5	heel ontevreden	ontevreden	niet slecht/ kan beter	tevreden	heel tevreden
Kolom 2	tevredenheid regelmaat OV	VAR 2	1 t/m 5	heel ontevreden	ontevreden	niet slecht/ kan beter	tevreden	heel tevreden
Kolom 3	tevredenheid bereik van OV	VAR 3	1 t/m 5	heel ontevreden	ontevreden	niet slecht/ kan beter	tevreden	heel tevreden
Kolom 4	tevredenheid horeca- faciliteiten	VAR 4	1 t/m 5	heel ontevreden	ontevreden	niet slecht/ kan beter	tevreden	heel tevreden
Kolom 5	tevredenheid onderwijs	VAR 5	1 t/m 5	heel ontevreden	ontevreden	niet slecht/ kan beter	tevreden	Heel tevreden
Kolom 6	tevredenheid werkgelegenheid	VAR 6	1 t/m 5	heel ontevreden	ontevreden	niet slecht/ kan beter	tevreden	heel tevreden
Kolom 7	tevredenheid huisvesting	VAR 7	1 t/m 5	heel ontevreden	ontevreden	niet slecht/ kan beter	tevreden	heel tevreden
Kolom 15	Respondent	VAR 8	Volgnr. 1 t/m 100					

Figuur 1. Aangepast overgenomen uit Voorbeeld codeboek (<https://www.scriptium.nl/codeboek/>)



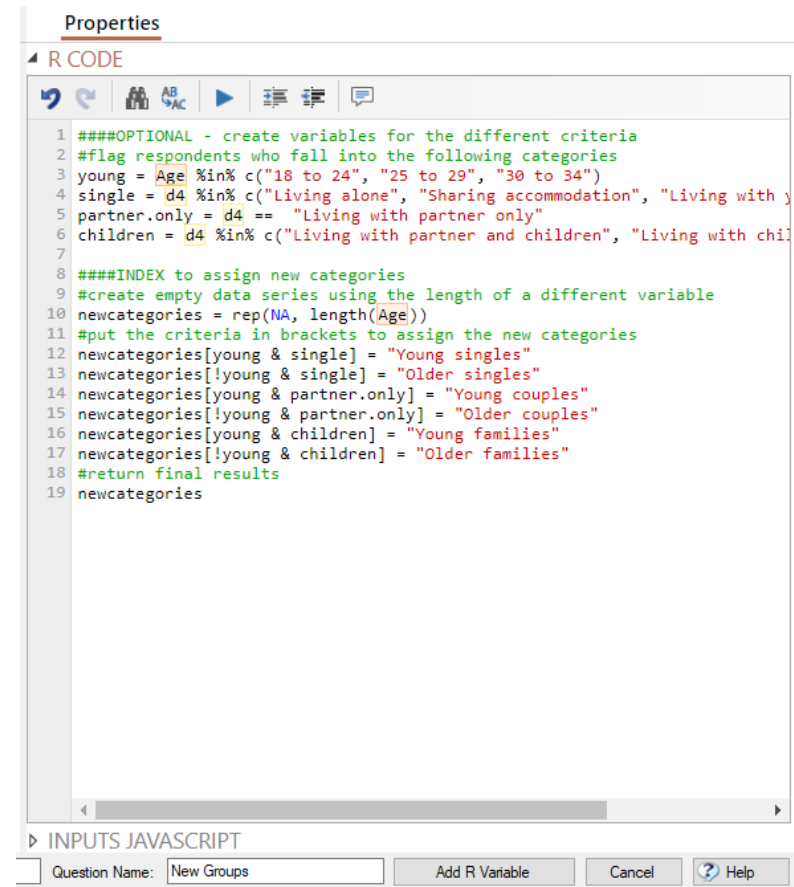
OMSCHRIJVING VARIABELEN SCRIPTS



```
main.py × +
1 #vraag de gebruiker naar zijn voornaam en
  familienaam
2 voornaam = input("Wat is je voornaam? ")
3 achternaam = input("Wat is je familienaam? ")
4
5 #combineer de voornaam en familienaam in de
  variabele volledige_naam
6 volledige_naam = voornaam + achternaam
7
8 #Toon de volledige naam op het scherm
9 print(volledige_naam)

Console × Shell × +
Wat is je voornaam? Jan
Wat is je familienaam? Peeters
JanPeeters
```

Figuur 3. Aangepast overgenomen uit Voorbeeld omschrijvingvariabelen python
(<https://informaticalessen.be/programmeren/de-functies-input-int-float-type-str-en-het-plakken-van-strings/string-aan-elkaar-plakken-python/>)



```
Properties
R CODE
1 #####OPTIONAL - create variables for the different criteria
2 #flag respondents who fall into the following categories
3 young = Age %in% c("18 to 24", "25 to 29", "30 to 34")
4 single = d4 %in% c("Living alone", "Sharing accommodation", "Living with y
5 partner.only = d4 == "Living with partner only"
6 children = d4 %in% c("Living with partner and children", "Living with chil
7
8 #####INDEX to assign new categories
9 #create empty data series using the length of a different variable
10 newcategories = rep(NA, length(Age))
11 #put the criteria in brackets to assign the new categories
12 newcategories[young & single] = "Young singles"
13 newcategories[!young & single] = "Older singles"
14 newcategories[young & partner.only] = "Young couples"
15 newcategories[!young & partner.only] = "Older couples"
16 newcategories[young & children] = "Young families"
17 newcategories[!young & children] = "Older families"
18 #return final results
19 newcategories

INPUTS JAVASCRIPT
Question Name: New Groups Add R Variable Cancel Help
```

Figuur 4. Aangepast overgenomen uit Voorbeeld omschrijvingvariabelen in R
(<https://help.qresearchsoftware.com/hc/en-us/articles/4413543602447-How-to-Create-a-New-Variable-Based-on-Other-Variables-using-R>)



OMSCHRIJVING VARIABELEN SPSS EN ATLAS.TI

*SAMPLE DATAMATRIX for KIT.sav [DataSet1] - SPSS Data Editor

	Name	Type	Width	Decimals	Label	Values	Miss
1	DSG	Numeric	8	0	Designation	{1, Lecturer}...	None
2	QUA	Numeric	8	0	Qualification	{1, Master}...	None
3	GDR	Numeric	8	0	Gender	{1, Male}...	None
4	EXP	Numeric	8	0	Length of service	None	None
5	Q5	Numeric	8	0		{1, SDA}...	None
6	Q6	Numeric	8	0		{1, SDA}...	None
7	Q7	Numeric	8	0		{1, SDA}...	None
8	Q8	Numeric	8	0		{1, SDA}...	None
9	Q9	Numeric	8	0		{1, SDA}...	None

Data View Variable View

SPSS Processor is ready

Figuur 5. Overgenomen uit 2017 Handbook for Social Researchers Prof Dr Allah Nawaz - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/variable-View-of-SPSS_fig11_336812698 [accessed 24 Jun, 2024]

Sentiment Analysis

Sentiment Analysis: Results
Review codings proposed by sentiment analysis, and add manual codings

524 paragraphs containing sentiments:
Positive
Neutral
Negative

Selected Documents (20)

- 17 Interview - Lay... 14
- 18 Interview - Lay... 20
- 19 Interview - Layp... 6
- 20 Interview - Layp... 9
- 21 Interview - Layp... 8
- 22 Interview - Layp... 6
- 23 Interview - Layp... 7
- 24 Interview - Layp... 8

17:16 ¶ 11 in Interview - Layperson 1
3 codings

As an adult who works full time and has a family, my main priority is to provide for my family and ensure that we have a comfortable and enjoyable life. While I understand the importance of sustainability, it is not a primary...

17:16 As an adult...
Attitude and mindset: Fa
Human values: Materiali
Sentiment: Positive

17:17 ¶ 13 in Interview - Layperson 1
3 codings

My focus is on making sure that my family has everything they need to be happy and healthy, and that includes the conveniences and luxuries that come with modern living. This means that I am willing to consume products and services t...

17:17 My focus is on...
Human values: Family
Human values: Materiali
Sentiment: Positive

17:18 ¶ 15 in Interview - Layperson 1
3 codings

I know that there are people who are deeply committed to living sustainably, and I respect their choices. However, for me, sustainability is not a top priority. I believe that there are trade-offs to be made in life, and I am willing to prior...

17:18 I know that th...
Adaptability: Balancing p
Environmental conscious
Sentiment: Positive

17:17 ¶ 17 in Interview - Layperson 1
No codings

Interviewer:
Do you do anything to live more sustainably at home, at work, or while on vacation?

Sentiment: Neutral

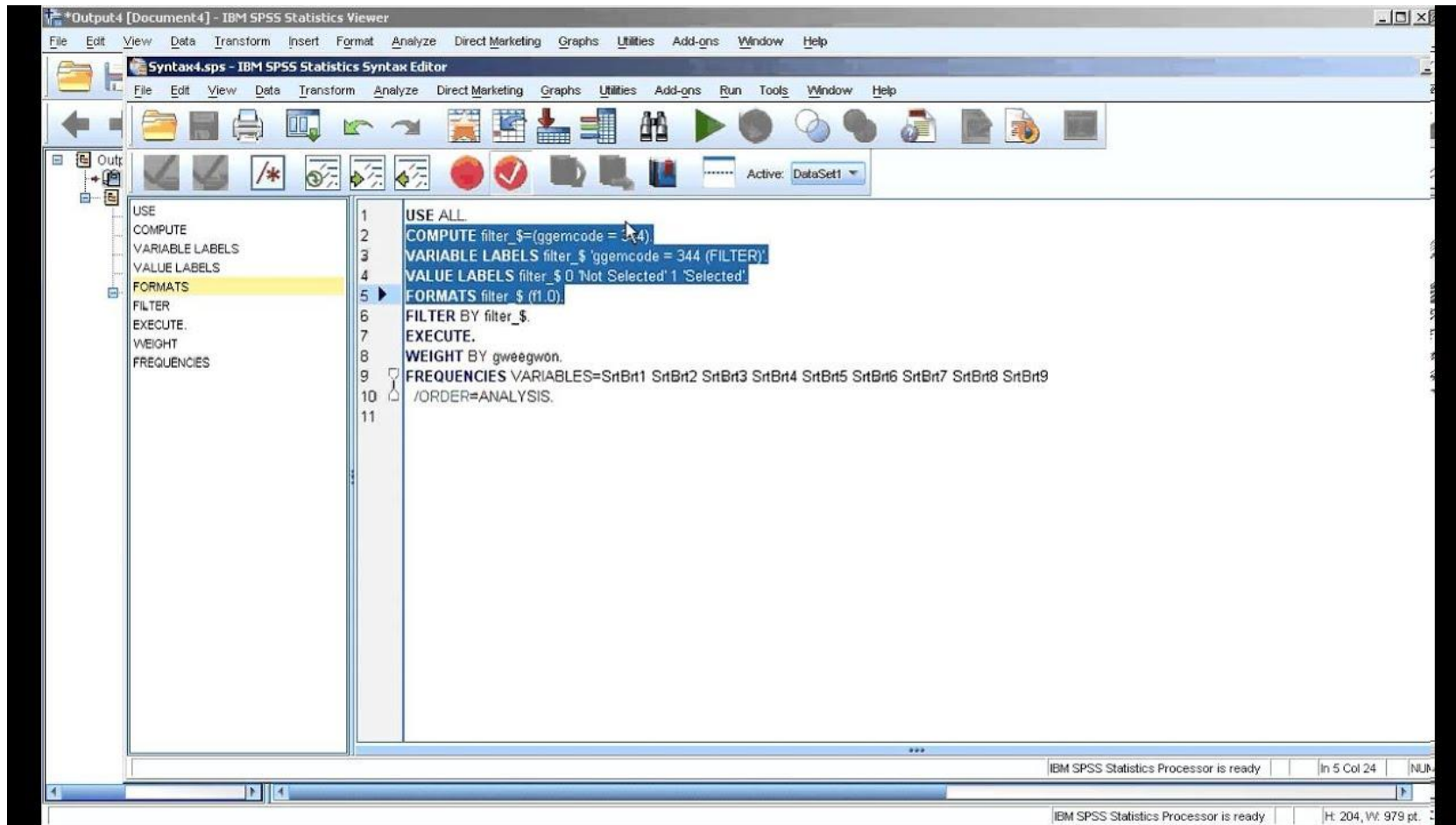
Back New Search

Figuur 6. Aangepast overgenomen uit Voorbeeld omschrijving variabelen in Atlas.TI (<https://atlasti.com/guides/qualitative-research-guide-part-2/qualitative-data-analysis#how-do-you-analyze-qualitative-data>)





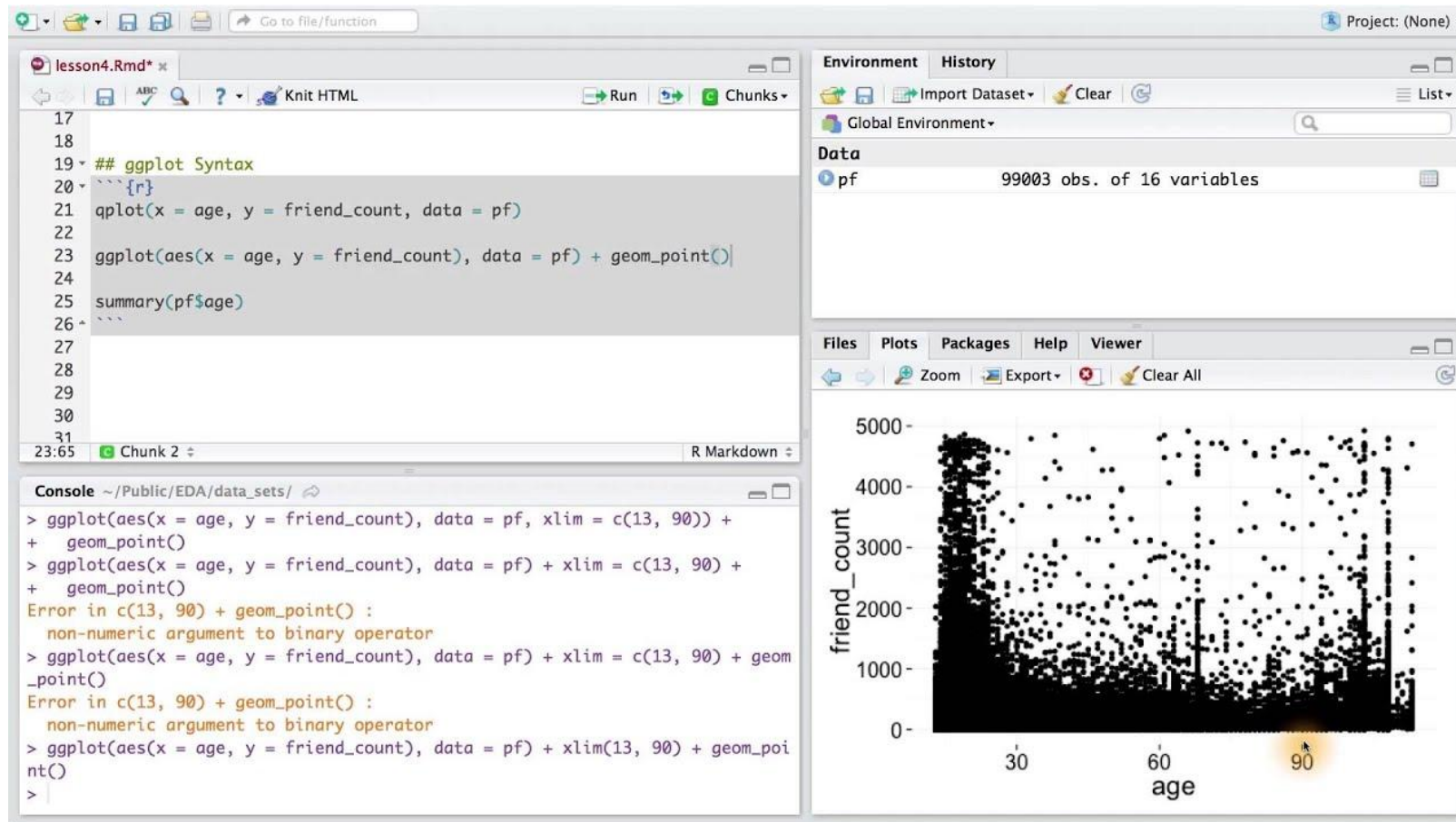
OMSCHRIJVING ANALYSE SPSS (SYNTAX)



Figuur 7. Aangepast overgenomen uit Voorbeeld omschrijving analyse SPSS (<https://www.youtube.com/watch?app=desktop&v=gwHISCdqz4Y>)



OMSCHRIJVING ANALYSE R



Figuur 8. Aangepast overgenomen uit Voorbeeld omschrijving analyse R (<https://www.youtube.com/watch?v=7XyFjPDPoZQ>)





OMSCHRIJVING ANALYSE PYTHON

```
import numpy as np

a = np.array([1,4,54,6,90,8.8,-3.2,67])
print(a)

# Compute and print various statistics
print('Mean:', a.mean())
print('Median:', np.median(a))
print('Range (Max - min):', np.ptp(a))
print('Standard deviation:', a.std())
print('80th percentile:', np.percentile(a, 80))
print('0.2-quantile:', np.quantile(a, 0.2))
```

Figuur 9. Aangepast overgenomen uit Voorbeeld 1 omschrijving analyse Python
(<https://m.dexlabanalytics.com/blog/python-statistics-fundamentals-how-to-describe-your-data-part-ii>)

```
import pingouin as pg
import pandas as pd
import numpy as np

#create DataFrame
df = pd.DataFrame({'score': [64, 66, 68, 75, 78, 94, 98, 79, 71, 80,
                             91, 92, 93, 90, 97, 94, 82, 88, 95, 96,
                             79, 78, 88, 94, 92, 85, 83, 85, 82, 81],
                  'group': np.repeat(['strat1', 'strat2', 'strat3'], repeats=10)})

#perform Welch's ANOVA
pg.welch_anova(dv='score', between='group', data=df)
```

Figuur 10. Aangepast overgenomen uit Voorbeeld 2 omschrijving analyse Python
(<https://www.geeksforgeeks.org/how-to-perform-welchs-anova-in-python/>)





Wordt er onderzoekssoftware gemaakt binnen het onderzoek? Denk aan een software management plan, zie [Practical guide to Software Management Plans](#).

Software publiceren kan via [Research Software Directory](#).





ONDERZOEKSPROTOCOL

Een onderzoeksprotocol bevat het volgende:

- introductie (onderzoeksvraag)
- methoden
 - design
 - werving deelnemers
 - dataverzameling
 - data-analyse
 - datamanagement
- ethische overwegingen
- toestemming proefpersoon
- valorisatie en publicatie
- financiering





INFORMED CONSENT

- voor deelname aan onderzoek is geïnformeerde toestemming (informed consent) nodig van de deelnemer
- in je datadocumentatie neem je een niet-ingevuld toestemmingsformulier (informed consent) op





INFORMATIEBRIEF

Met een informatiebrief worden deelnemers voor het onderzoek geïnformeerd over:

- het doel van het onderzoek
- de opdrachtgever
- wie het onderzoek uitvoert
- wat er van de deelnemers verwacht wordt (welke handelingen verricht worden/welke vragenlijsten ingevuld moeten worden etc.)
- rechten van de deelnemer
- wat de voor- en nadelen van deelname aan het onderzoek zijn
- waar eventueel meer informatie over het onderzoek op te vragen is





VRAGENLIJST

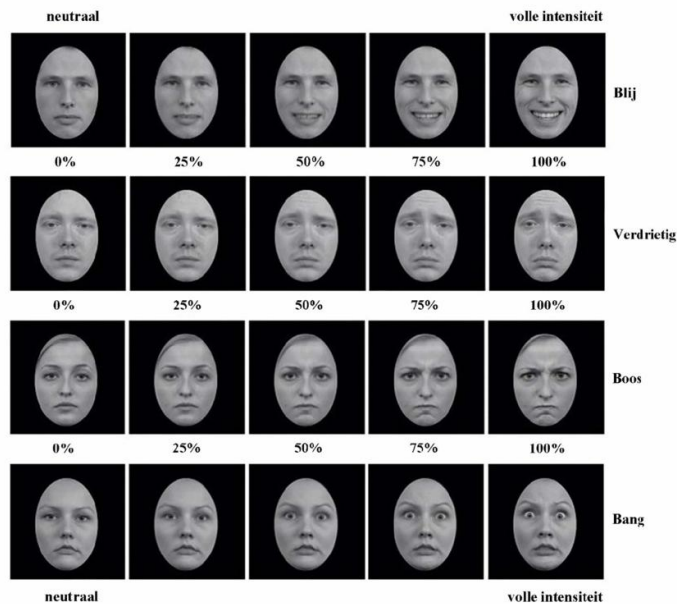
In je datadocumentatie neem je (een) niet-ingevulde vragenlijst(en) op.

Voorbeelden zijn:

- digitale vragenlijst
- vragen die je tijdens interview wil stellen
- overzicht met onderwerpen die je tijdens een interview of met een focusgroep wil bespreken
- observatieschema



Alles wat wordt aangeboden aan een deelnemer in een experimentele conditie om een bepaalde response te ontlokken.



Figuur 11. Aangepast overgenomen uit *Voorbeeld 1 Stimuli*
(https://www.researchgate.net/figure/Figuur-1-Voorbeeld-van-stimuli-uit-de-dynamische-gezichtsemotieherkenningstaak-waarin_fig1_338588344)

Demonstratie Stroop-effect
Noem zo snel mogelijk de kleuren van de woorden.

Groen	Blauw
Blauw	Rood
Paars	Groen
Blauw	Rood
Rood	Paars

Congruent **Incongruent**

Het Stroop-effect houdt in dat het meer tijd kost om de kleuren te benoemen in de incongruente dan in de congruente conditie.

Figuur 12. Aangepast overgenomen uit *Voorbeeld 2 Stimuli*
(<https://nl.wikipedia.org/wiki/Stroop-taak>)





VOCABULAIRE

Vocabulaires zijn woordenlijsten die gebruikt kunnen worden voor labelen, indexeren of categoriseren. Vocabulaires zorgen voor consistentie en interoperabiliteit en zijn machine leesbaar. Er zijn verschillende soorten:

- [gecontroleerde vocabulaire](#): woordenlijst. Beleid bepaalt wie, wanneer en hoe woorden aan de lijst worden toegevoegd. Alleen woorden uit de lijst mogen gebruikt worden.
- [taxonomie](#): classificatieschema, vaak in een boomstructuur. Entiteiten worden beschreven onderverdeeld en geordend in hiërarchische groepen en categorieën. Bevat criteria waaraan moet worden voldaan om tot de categorie te horen. Komt tot stand op basis van metingen en experimenten.
- [thesaurus](#): gestructureerde gecontroleerde woordenlijst, waarin hiërarchische relaties worden gelegd naast gerelateerde termen en synoniemen.
- [ontologie](#): woordenlijst, bevat abstracte en concrete concepten maar ook processen en taken. Tussen de concepten worden relaties aangegeven die zorgen voor structuur en ordening. Bevat ook de regels waaraan de concepten en relaties binnen een bepaald vakgebied moeten voldoen.
- [typologie](#): een conceptuele classificatie (ontologie) op basis van logisch redeneren.





README-FILE

Een README-file kan voor het hele onderzoeksproject geschreven worden, maar als de dataset complex is ook per werkpakket of bestand.

Het bevat bijvoorbeeld:

- omschrijving van de mappenstructuur
- omschrijving van de totstandkoming van de benaming
- inhoud van de files en hun format
- relatie met andere datasets
- methodologische informatie
- informatie over de rechten en het al dan niet kunnen delen van (vertrouwelijke) data
- je kunt een README-file maken met het Windows Kladblok of met een MAC TextEdit (kies "Make plain tekst"). Meer complexe README-files zijn bv Markdown codeboeken
- Een voorbeeld template voor een README-file





METADATASTANDAARD

Metadatastandaarden zijn vastgestelde schema's van metadata bestaande uit verschillende elementen met informatie. Vaak heeft een repository al vastgesteld welke metadatastandaard je moet gebruiken.

- DataverseNL gebruikt bijvoorbeeld de Dublin Core Metadata Standard, bestaande uit 15 elementen [DCMI_15 elements](#)
- als er geen metadatastandaard gespecificeerd is in de geselecteerde repository, kun je zelf zoeken naar een geschikte metadatastandaard [Guidance metadata standard](#)





PERSISTENT IDENTIFIER

Een Persistent Identifier (PID of PI) is een permanente verwijzing en uniek label naar een digitale bron. De digitale bron kan een persoon, plaats of ding (bv. artikel of dataset) zijn. Het doel van een PID is dat met deze code de digitale bron altijd vindbaar blijft, ook als de URL (het webadres) van de digitale bron of de naam wordt gewijzigd. Een PID is dus een stabiele link naar een digitale bron.

Voorbeelden van een PID zijn:

- ORCID (Open Researcher and Contributor ID) is een PID voor onderzoekers. Om te waarborgen dat je publicaties daadwerkelijk aan jou gelinkt worden, is het van belang dat je met jouw ORCID in de onderzoekersdatabase van de Repository geregistreerd staat. Je kunt je ORCID aanvragen op de website van ORCID.
- DOI (Digital Object Identifier) is een PID voor een dataset en wordt bij deponeren in een repository automatisch toegekend. Dit hoef je dus niet zelf aan te vragen.
- RAiD (Research Activity Identifier) is een PID voor een onderzoeksproject
- ROR (Research Organization Registry) is een PID voor een onderzoeksinstituting





LICENTIES

Door je onderzoeksdata te voorzien van een licentie bepaal je vooraf in welke mate de onderzoeksdata hergebruikt mag worden door anderen.

- Creative Commons wordt vaak gebruikt voor data uit kwantitatief en kwalitatief onderzoek. Er zijn verschillende [CC licenties](#) waar je uit kunt kiezen
- voorbeelden van licenties voor software en code vind je bij het [Open Source Initiative](#)





BESTANDSFORMATEN

Voorkeursformaten zijn de bestandsformaten die te lezen zijn in vrij verkrijgbare software, waarvan goede en open documentatie beschikbaar is, en die veel gebruikt worden (binnen een bepaalde discipline).

- door te kiezen voor voorkeursformaten maak je je onderzoeksdata en datadocumentatie herbruikbaar, toegankelijk en duurzaam. Bestanden zijn dan over een aantal jaren nog te openen.
- repositories stimuleren het gebruik van voorkeursformaten door een overzicht te bieden, zoals [4TU.ResearchData](#) en [DANS](#). Repositories verschillen in hun aanbeveling.
- repositories maken meestal onderscheid tussen voorkeursformaten (preferred of recommended) en acceptabele formaten (non-preferred formats of acceptable formats).





MAPPENSTRUCTUUR

Maak je bestanden vindbaar door op het volgende te letten:

- laat de fase of type van onderzoek terugkomen in de mappenstructuur
- bewaar ruwe data in een aparte map
- maak niet te veel niveaus in je mappenstructuur
- documenteer de mappenstructuur in een [README-file](#)
- hou met de opzet van je mappenstructuur rekening met het delen van data





BESTANDSNAMEN

Maak je bestanden vindbaar door op het volgende te letten (zie ook [RDNL](#))

- geef het bestand een betekenisvolle naam (Project1_Manuscript)
- hou een consistente benaming aan (Project1_Manuscript
Project2_Manuscript_YYYYMMDD)
- gebruik geen spaties (maar _ of -) en geen speciale tekens
- benoem het versie nummer (Project1_Data_V1, Project1_Data_V2)
- maak de benaming niet te lang

