

WAT VERKLAART DE NSE?

METHODE EN AANPAK VU STUDENT ANALYTICS
& ILLUSTRATIE AAN DE HAND VAN DE NSE



DAIR

2 NOVEMBER 2017



WAT VERKLAART DE NSE?

1. Wat verklaart de NSE?

- a) Relevantie voor de VU
- b) Methode / Variabelenselectie
- c) Bevindingen

2. Onze kwaliteitscyclus

- a) De statistische waardeketen
- b) Coördinatie en samenwerking
- c) Herhaalbaarheid
- d) Borging binnen de VU

Relevantie voor de VU

- Landelijke vragenlijst → benchmark
- Hoge respons: 7.000-10.000 studenten (31,7%)
- Hybride doel: voorlichting en kwaliteitszorg
- Brede scope



Komt voor in:

- Managementrapportages
- Jaarverslagen
- Besprekingen met studenten
- Themabijeenkomsten
opleidingsdirecteuren en bestuurders

[illegible]



Data exploratie en modellering

-  1. Variabelenselectie Y en X
-  2. Dataexploratie:
 - Outliers (uitbijters)
 - Mogelijke interacties
 - Correlaties tussen X-variabelen (collineariteit)
-  3. Modelleren
-  4. Modelvalidatie
-  5. Bevindingen

Gebaseerd op Zuur et al. (2009), *A protocol for data exploration to avoid common statistical problems*



Data
Scientist



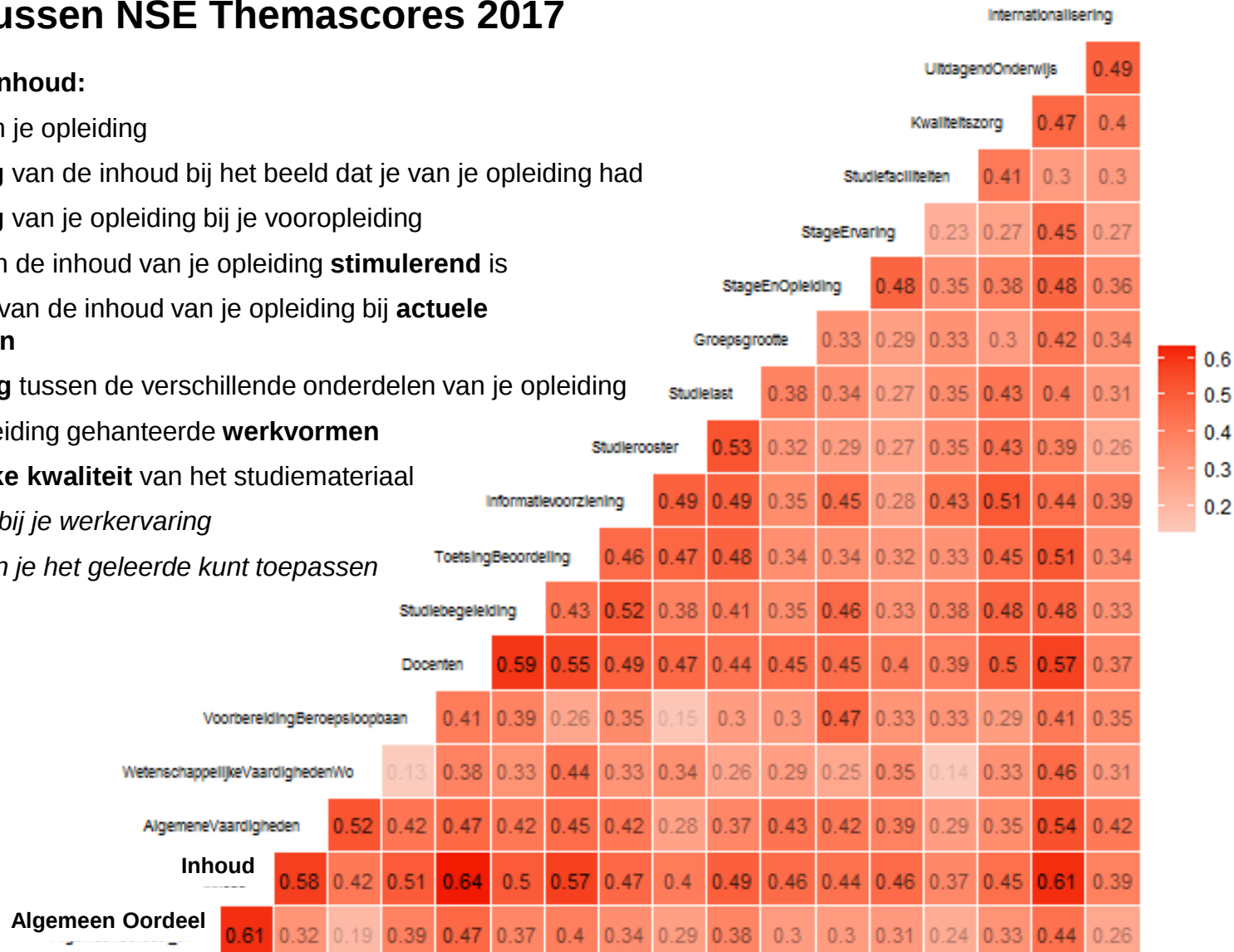
Correlaties tussen NSE Themascoringen 2017

NSE Themascoring Inhoud:

- Het **niveau** van je opleiding
- De **aansluiting** van de inhoud bij het beeld dat je van je opleiding had
- De **aansluiting** van je opleiding bij je vooropleiding
- De mate waarin de inhoud van je opleiding **stimulerend** is
- De aansluiting van de inhoud van je opleiding bij **actuele ontwikkelingen**
- De **samenhang** tussen de verschillende onderdelen van je opleiding
- De in jouw opleiding gehanteerde **werkvormen**
- De **inhoudelijke kwaliteit** van het studiemateriaal
- *De aansluiting bij je werkervaring*
- *De mate waarin je het geleerde kunt toepassen in je baan*



Data Scientist





Model:

$$Y \sim X_1 + X_2 + X_3 + \dots + X_n,$$

Waarbij: Y = NSE Themascoring Inhoud

X_i = verklarende variabelen



Bron: Analyseset Student Analytics

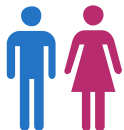


**Data
Scientist**



Dataset Student Analytics

Als basis voor deze analyse is de dataset van Student Analytics gebruikt. Hierin zijn veel verschillende gegevens verzameld voor analyses, beleidsontwikkeling en wetenschappelijk onderzoek waarvan het grootste gedeelte wordt gebruikt voor analyses van studiesucces:



Geslacht
Leeftijd op 1 oktober
Land van herkomst VO
Gezinssamenstelling

Demografie

Vooropleiding & aansluiting

Profielkeuze
Vooropleiding en onderwijsinstelling
Tussenjaren / Jaren sinds diploma
Verblijfsjaren

Voorlichting en introductie
Matching

Stroominformatie
Geografische spreiding (GIS)
Kwaliteit scholen
Achtergrond scholen

Instroom

Studiesucces & tevredenheid

Eindexamencijfers
Uitslagen taaltoets VU
Studieprestaties aan de VU
Honours en Cum Laude
Studenttevredenheid
Uitval en diplomarendement



Data
Scientist

De **rode variabelen** zijn geselecteerd op basis van eerder onderzoek naar studiesucces. Het vermoeden is dat studenttevredenheid en studiesucces met elkaar samenhangen.



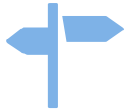


Het doel is een **transparant model**. Het is duidelijk uit te leggen welke variabelen bijdragen en in welke mate



Persoonskenmerken

Geslacht
Leeftijd op 1 oktober



Voorlichting gevolgd



Vooropleiding



Matching

Aantal uren nevenactiviteiten
Vertrouwen in studiesucces
Proactief studeergedrag
Concentratievermogen
Behoeftte aan informatie
Score voor ingeschreven opleiding
Ingeschreven voor hoogste interesse
Hulp van omgeving
Ondersteuning nodig



**Data
Scientist**



Studieomstandigheden (NSE)

Aantal uren werk en studie



Opleiding

Studiejaar



Aantal studies

Dubbele studie



Resultaten

Totaalresultaat taaltoets
Aantal EC in jaar 1
Gemiddeld resultaat in jaar 1



Hypothesen

Verwachte positieve relaties:

Geslacht_vrouw	+
Voorlichting gevolgd	+
Hoogste vooropleiding_vwo	+
Score voor ingeschreven opleiding	+
Ingeschreven voor hoogste interesse	+
Totaalresultaat taaltoets	+
Aantal EC in jaar 1	+
Gemiddeld resultaat in jaar 1	+
Dubbele studie	+
Hulp van omgeving	+
Vertrouwen in studiesucces	+
Proactief studeergedrag	+
Concentratievermogen	+

Verwachte negatieve relaties:

Leeftijd op 1 oktober	-
Aantal uren nevenactiviteiten	-
Behoeftte aan informatie	-
Ondersteuning nodig	-
Aantal uren werk	-
Aantal uren studie	-
Studiejaar	-



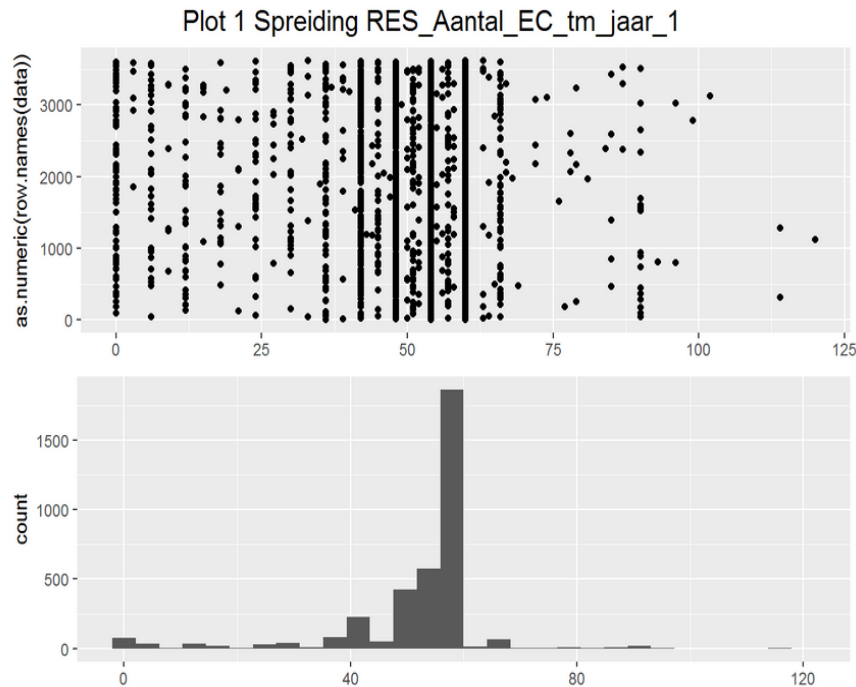


Dataselectie vooraf

Omdat we benieuwd zijn naar een verklaring voor de NSE-score van de reguliere studenten, zijn er vooraf de volgende selecties gemaakt:

- NSE 2017, dit is collegejaar 2016-2017
- NSE ingevuld
- Bachelor
- Vooropleiding = vwo of hbo-p
- Max. studiejaar = nominaal + 1





- Leeftijd op 1 oktober > 30 jr
- MVR uur/wk nevenactiviteiten > 40
- MVR Beeldvorming < 20
- MVR vertrouwen studiesucces < 10
- Aantal uur werk per week > 20
- Aantal EC in jaar 1 > 80

Na het verwijderen van uitbijters zijn er nog 3439 observaties over voor de analyse

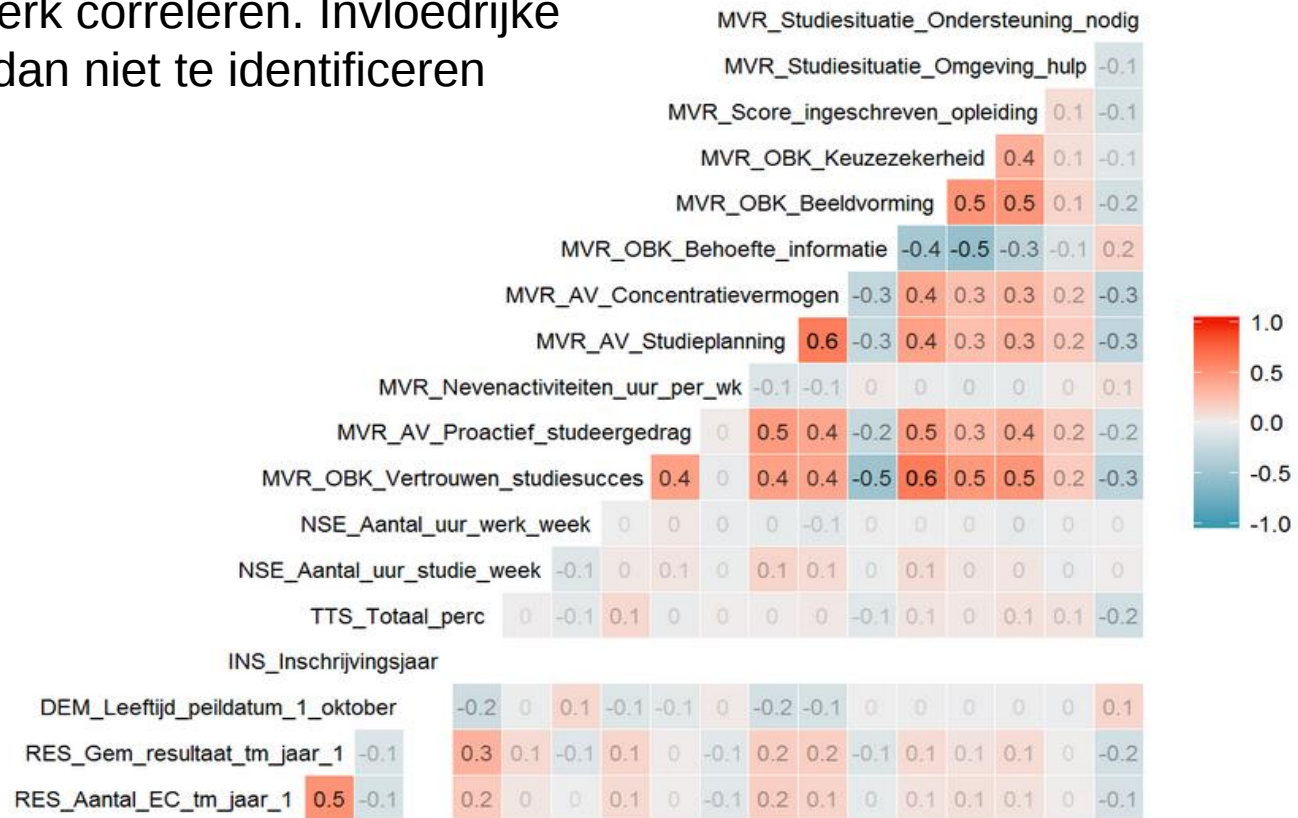


**Data
Scientist**

key	0	1	NA_
<chr>	<dbl>	<dbl>	<dbl>
OUT_INS_Leeftijd_hoger_dan_30	3604	9	0
OUT_MVR_Nevenactiviteiten_uur_per_wk_boven_40	2973	13	627
OUT_MVR_OBK_Beeldvorming_onder_25	2995	1	617
OUT_MVR_OBK_Vertrouwen_studiesucces_onder_10	2989	7	617
OUT_NSE_meer_dan_20_uur_werk	2975	112	526
OUT_RES_Meer_dan_80_EC	3577	36	0
6 rows			



Collineariteit treedt op wanneer X-variabelen te sterk correleren. Invloedrijke variabelen zijn dan niet te identificeren met het model.



Data
Scientist

Er is veel samenhang tussen de verschillende variabelen uit de matchingsvragenlijst en tussen de resultaten in het eerste jaar
→ Risico op collineariteit!



VIF-test

Met behulp van een VIF-test (*Variance Inflation Factor*) zijn de volgende variabelen stapsgewijs verwijderd uit de selectie:

- MVR_OBK_Beeldvorming
- MVR_AV_Studieplanning
- MVR_OBK_Vertrouwen_studiesucces
- MVR_OBK_Keuzezekerheid



Data
Scientist

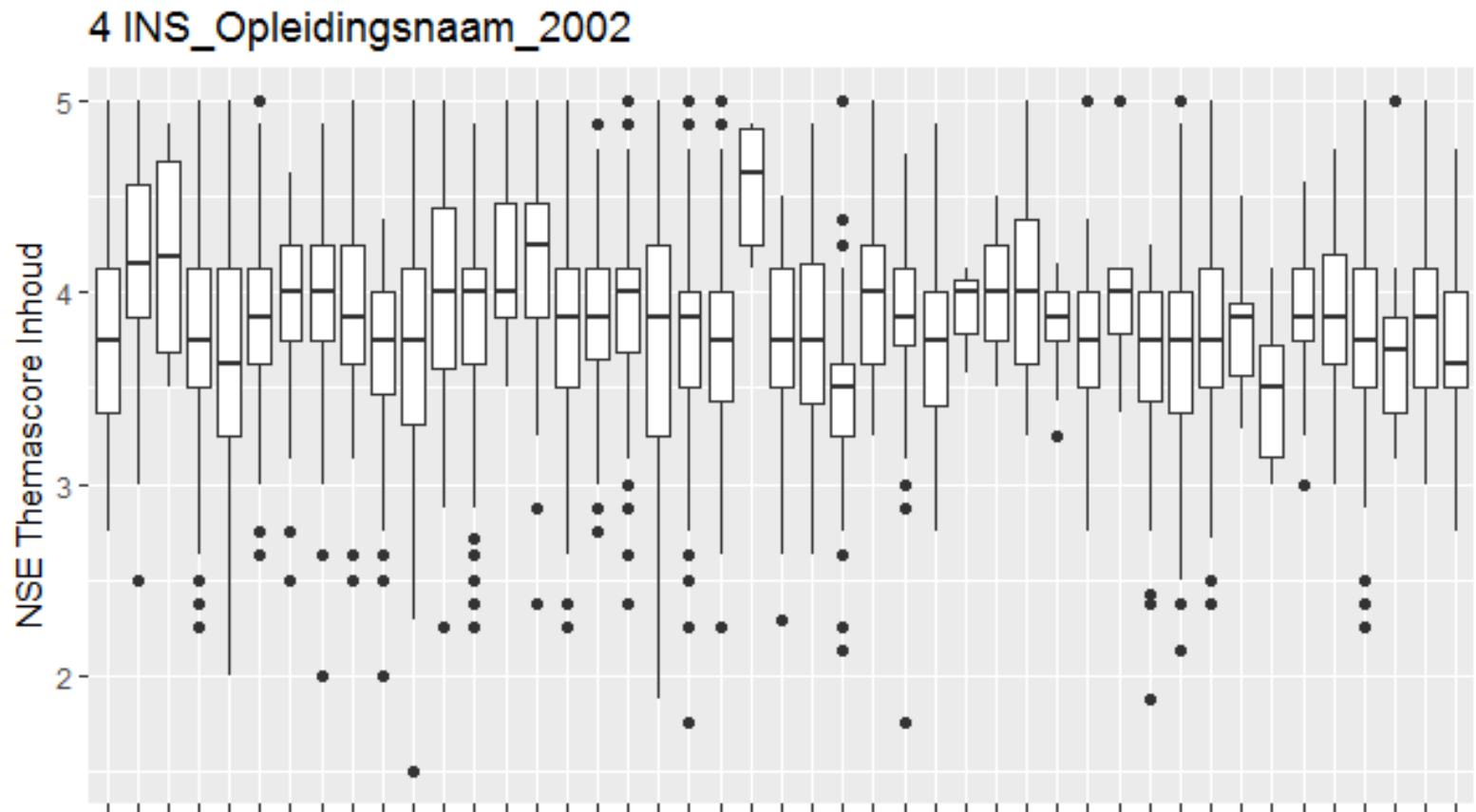
Call:

```
mctest::imcdiag(x = x, y = data[, y_var], method = "VIF", vif = 2)
```

VIF Multicollinearity Diagnostics

	VIF detection	
RES_Aantal_EC_tm_jaar_1	1.2464	0
RES_Gem_resultaat_tm_jaar_1	1.3896	0
DEM_Leeftijd_peildatum_1_oktober	1.1279	0
INS_Inschrijvingsjaar	2.0000	1
TTS_Totaal_perc	1.2221	0
NSE_Aantal_uur_studie_week	1.0605	0
NSE_Aantal_uur_werk_week	1.0505	0
MVR_OBK_Vertrouwen_studiesucces	2.3301	1
MVR_AV_Proactief_studeergedrag	1.5861	0
MVR_Nevenactiviteiten_uur_per_wk	1.0352	0
MVR_AV_Studieplanning	2.1413	1
MVR_AV_Concentratievermogen	1.8859	0
MVR_OBK_Behoefte_informatie	1.6005	0
MVR_OBK_Beeldvorming	2.1942	1
MVR_OBK_Keuzezekerheid	1.7253	0
MVR_Score_ingeschreven_opleiding	1.5138	0
MVR_Studiesituatie_Omgeving_hulp	1.0801	0
MVR_Studiesituatie_Ondersteuning_nodig	1.2209	0

Multicollinearity may be due to INS_Inschrijvingsjaar MVR_OBK_Vertrouwen
VR_OBK_Beeldvorming regressors



Data
Scientist



Lineair mixed model

Met: random = ~ (RES_Aantal_EC_tm_jaar_1 | INS_Opleidingsnaam_fix)

Dit model houdt rekening met de variatie tussen opleidingen, en door een *random slope* toe te voegen, houdt het er ook rekening mee dat het verband tussen het aantal EC in B1 en de NSE-score verschilt per opleiding.

Interacties

INS_Hoogste_vooropleiding_soort_cat * MVR_Studiesituatie_Ouders_universiteit
INS_Dubbele_studie_VU * RES_Gem_resultaat_tm_jaar_1



Data
Scientist

```
Family: gaussian  
Link function: identity
```

```
Formula:
```

```
NSE_Themascore_Inhoud ~ RES_Aantal_EC_tm_jaar_1 + RES_Gem_resultaat_tm_jaar_1 +  
  DEM_Leeftijd_peildatum_1_oktober + TTS_Totaal_perc + NSE_Aantal_uur_studie_week +  
  NSE_Aantal_uur_werk_week + MVR_AV_Proactief_studeergedrag +  
  MVR_Nevenactiviteiten_uur_per_wk + MVR_AV_Concentratievermogen +  
  MVR_OBK_Behoefte_informatie + MVR_Score_ingeschreven_opleiding +  
  MVR_Studiesituatie_Omgeving_hulp + MVR_Studiesituatie_Ondersteuning_nodig +  
  INS_Studiejaar + INS_Dubbele_studie_VU + INS_Vooropleiding +  
  DEM_Geslacht + ORI_Orientatie_komt_voor + MVR_Ingeschreven_voor_hoogste_interesse +  
  MVR_Ouders_universiteit + INS_Vooropleiding * MVR_Ouders_universiteit +  
  INS_Dubbele_studie_VU * RES_Gem_resultaat_tm_jaar_1
```



Parametric coefficients:

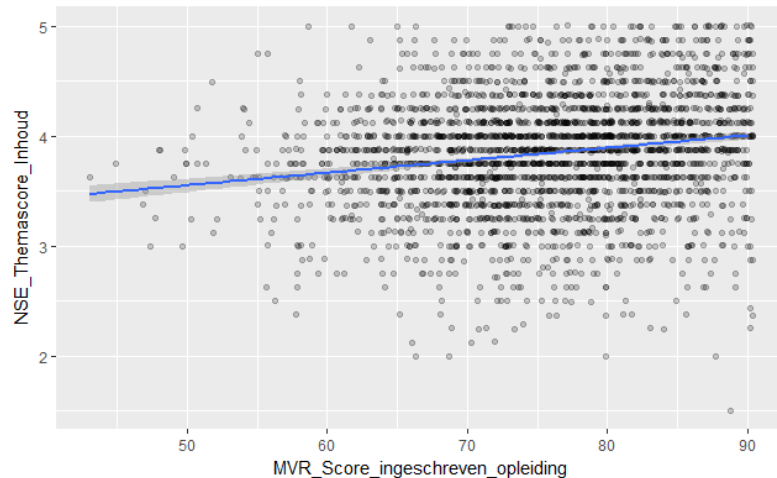
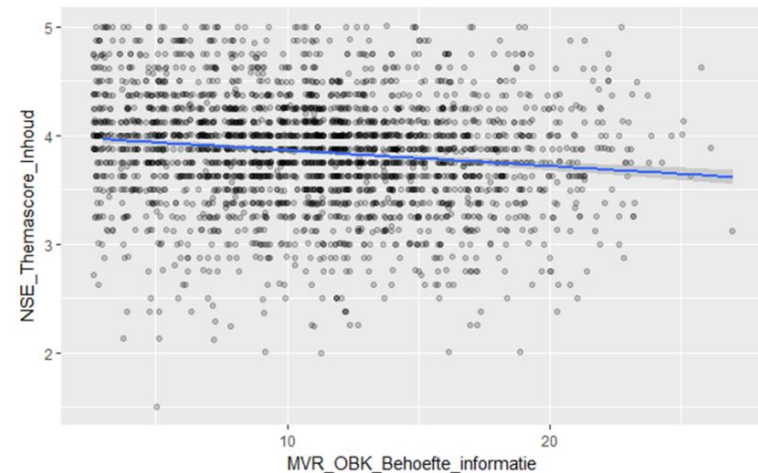
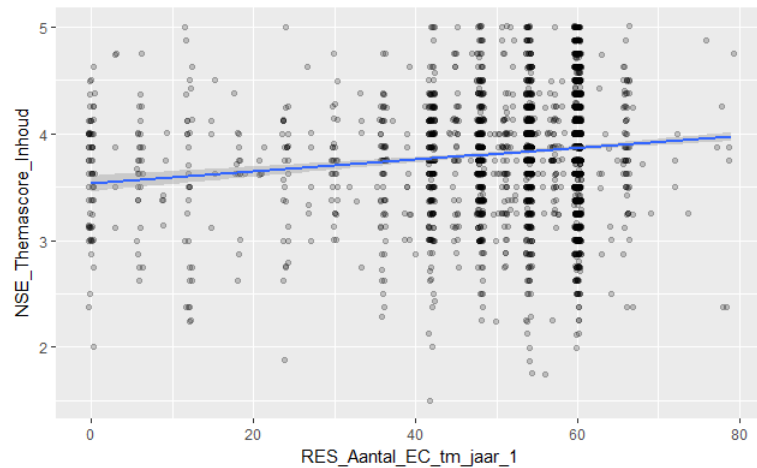
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.4178458	0.3083200	7.842	6.86e-15	***
RES_Aantal_EC_tm_jaar_1	0.0062799	0.0016981	3.698	0.000222	***
RES_Gem_resultaat_tm_jaar_1	0.0360782	0.0211525	1.706	0.088219	.
DEM_Leeftijd_peildatum_1_oktober	0.0079259	0.0082675	0.959	0.337823	
TTS_Totaal_perc	0.0009447	0.0017663	0.535	0.592806	
NSE_Aantal_uur_studie_week	0.0019638	0.0008839	2.222	0.026403	*
NSE_Aantal_uur_werk_week	-0.0009147	0.0018510	-0.494	0.621244	
MVR_AV_Proactief_studeergedrag	0.0004362	0.0027510	0.159	0.874019	
MVR_Nevenactiviteiten_uur_per_wk	-0.0002825	0.0017520	-0.161	0.871936	
MVR_AV_Concentratievermogen	0.0019043	0.0022888	0.832	0.405506	
MVR_OBK_Behoefte_informatie	-0.0099232	0.0022925	-4.329	1.57e-05	***
MVR_Score_ingeschreven_opleiding	0.0080110	0.0014893	5.379	8.28e-08	***
MVR_Studiesituatie_Omgeving_hulp	0.0181482	0.0095305	1.904	0.057012	.
MVR_Studiesituatie_Ondersteuning_nodig	0.0348182	0.0139297	2.500	0.012507	*
INS_Studiejaar2	-0.0556151	0.0260258	-2.137	0.032714	*
INS_Studiejaar3	-0.1463511	0.0296115	-4.942	8.30e-07	***
INS_Dubbele_studie_VUTRUE	1.0277192	0.5136779	2.001	0.045547	*
INS_VooropleidingVWO	-0.0619417	0.0478483	-1.295	0.195614	
DEM_Geslachtvrouw	-0.0249615	0.0246518	-1.013	0.311381	
ORI_Orientatie_komt_voorTRUE	0.0173449	0.0324220	0.535	0.592722	
MVR_Ingeschreven_voor_hoogste_interesse1	-0.0195842	0.0234527	-0.835	0.403779	
MVR_Ouders_universiteit1	-0.2714473	0.0761240	-3.566	0.000370	***
INS_VooropleidingVWO:MVR_Ouders_universiteit1	0.2692132	0.0790914	3.404	0.000676	***
RES_Gem_resultaat_tm_jaar_1:INS_Dubbele_studie_VUTRUE	-0.1630414	0.0700406	-2.328	0.020012	*

 signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

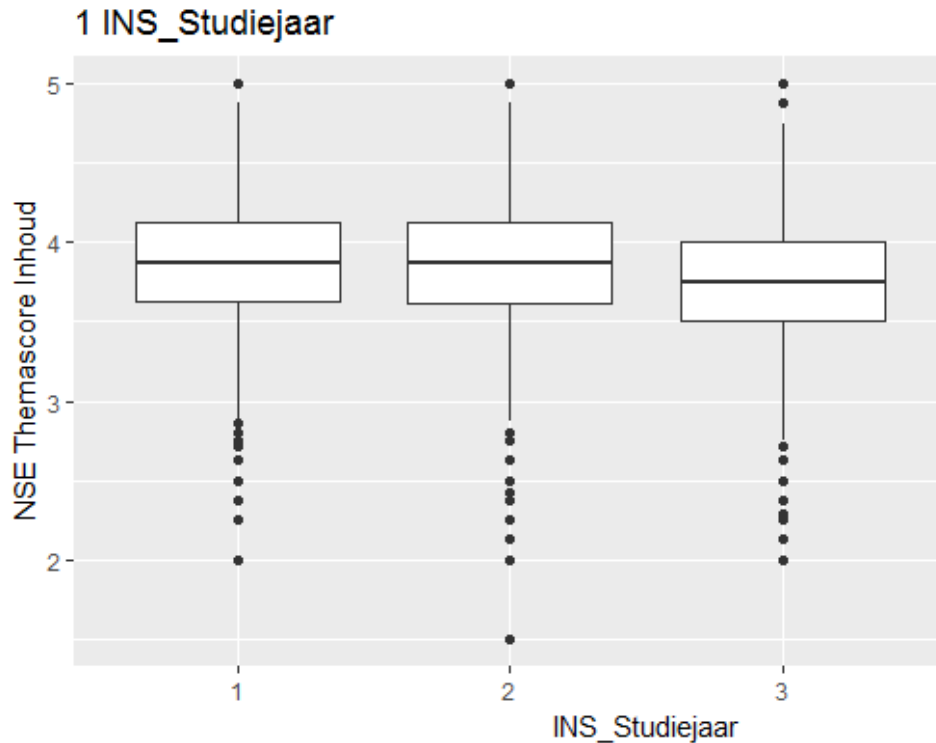
R-sq. (adj) = 0.0867
 lmer.REML = 3183.8 Scale est. = 0.2212 n = 2222

Dit model kan 9% van de variantie in de NSE themascore Inhoud verklaren.





- Succesvollere studenten in B1 geven hogere NSE-scores
- Studenten met een hoge interessescore voor hun opleiding geven hogere NSE-scores
- Studenten met meer behoefte aan informatie over de opleiding geven lagere NSE-scores

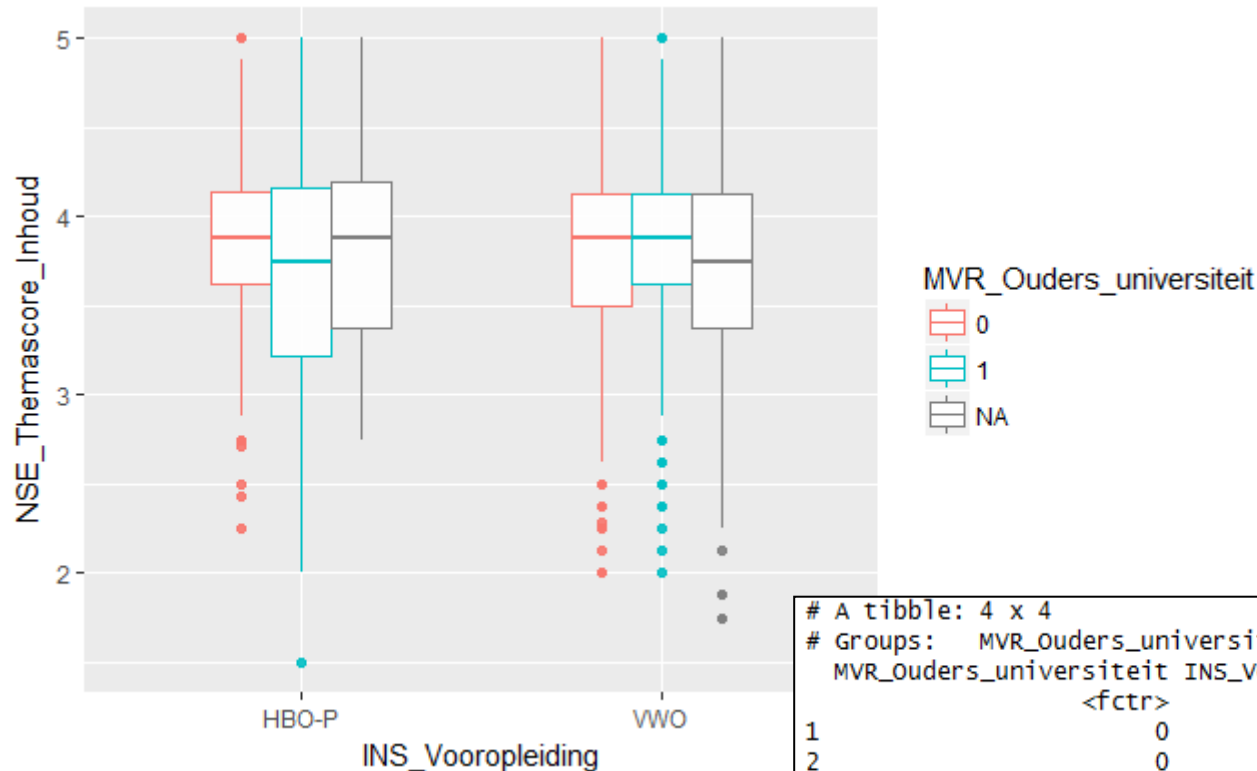


INS_Studiejaar	meanNSE	sdNSE
1	3.884102	0.4952467
2	3.879407	0.5101399
3	3.772953	0.5179070



Data
Scientist

Derdejaarsstudenten zijn minder tevreden ($p < 0,001$)

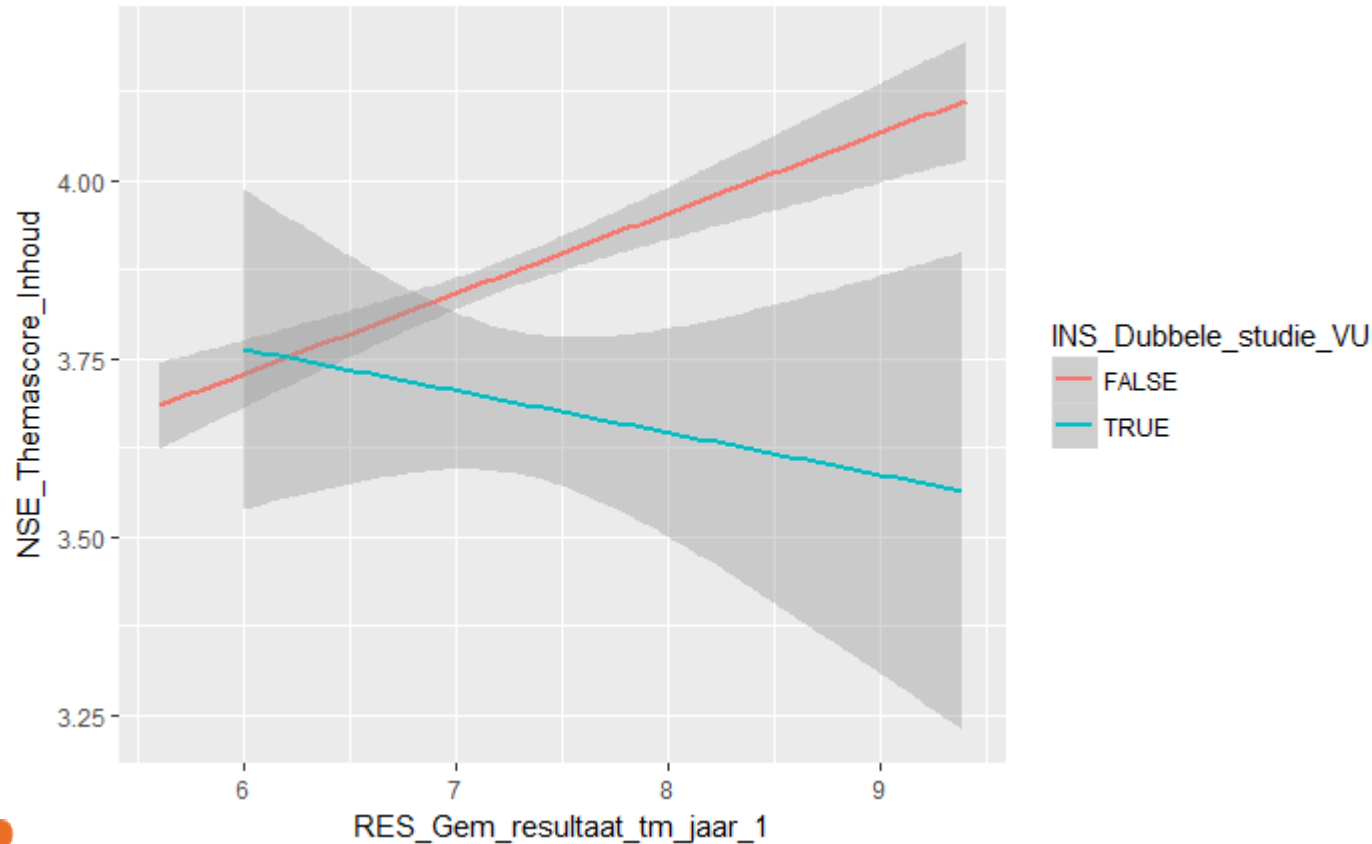


```
# A tibble: 4 x 4
# Groups:   MVR_Ouders_universiteit [?]
  MVR_Ouders_universiteit INS_vooropleiding meanNSE    sdNSE
      <fctr>                <fctr>      <dbl>    <dbl>
1                0          HBO-P  3.884627 0.5378502
2                0          VWO   3.850309 0.4974196
3                1          HBO-P  3.639724 0.7559826
4                1          VWO   3.860065 0.4963434
```



**Data
Scientist**

Studenten met een hbo-p vooropleiding, wiens ouders aan een universiteit hebben gestudeerd zijn minder tevreden dan de overige groepen. ($p < 0,001$)



Data
Scientist

Studenten met een dubbele studie, zijn kritischer naarmate zij gemiddeld hogere cijfers halen, terwijl de reguliere studenten tevredener zijn naarmate zij hogere cijfers halen. ($p < 0,05$)



1. Hoe presteert het model?
 - Root Mean Squared Error
 - Mean Absolute Error

$$RMSE = \sqrt{\frac{\sum (NSEscore - fit)^2}{N}}$$

$$MAE = \frac{\sum |NSEscore - fit|}{N}$$

2. Hoe presteert het model in vergelijking met het Nulmodel?
Nul-model = gemiddelde NSE-score per opleiding



Data
Scientist

	RMSE	MAE
Model	0,48	0,37
Nulmodel	0,51	0,39



Kracht

Het model voorspelt 9% van de variantie

Het model voorspelt iets beter dan het nulmodel

Hypothesen

Positieve effecten:

- Aantal EC in B1
- Hoge interessescore voor hun opleiding in matchingsvragenlijst

Negatieve effecten:

- Hoge behoefte aan informatie over de opleiding
- Derdejaarsstudent

Geen effecten, maar wel verwacht:

- Geslacht
- Gemiddeld studieresultaat in B1
- Deelname aan oriëntatie
- Leeftijd
- Aantal uur werk per week
- Aantal uur nevenactiviteiten per week
- Ingeschreven voor studie met hoogste interessescore

VERWERKING EN COMBINATIE VAN DATA BINNEN STUDENT ANALYTICS

Brondata is verzameld, gecombineerd, opgeschoond en getransformeerd en omgezet naar nieuwe datasets en rapportages

Brondata

Databestanden zijn gepseudonimiseerd en op een beveiligde netwerkschijf gezet.

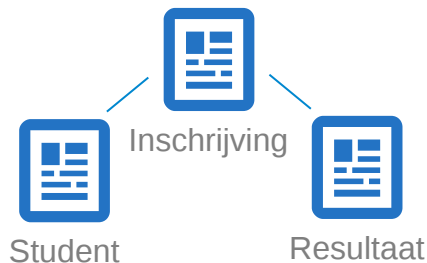


Bronnen

- DUO
- CBS
- MIVU / SAP SLM
- Selligent
- NOA (Matching)
- Introductie
- Taaltoets & bijspijker
- NSE
- Alumni

Datamodel & datamanipulatie

De data is vervolgens in een model gevat en verder bewerkt*



Modellering

- Tabellen per sleutel/niveau

Manipulatie:

- Transformeren, schonen, filteren
- Start met inschrijvingstabel, toevoegen van additionele informatie.

**Zie voor een toelichting op selecties en keuzes bijlage X*

Resultaten

Vervolgens zijn 4 views op de data gemaakt voor verschillende doeleinden*



Analysesets

Gecombineerde data, één regel per inschrijving. Resultaten met één regel per vak

Verrijkt 1CHO bestand

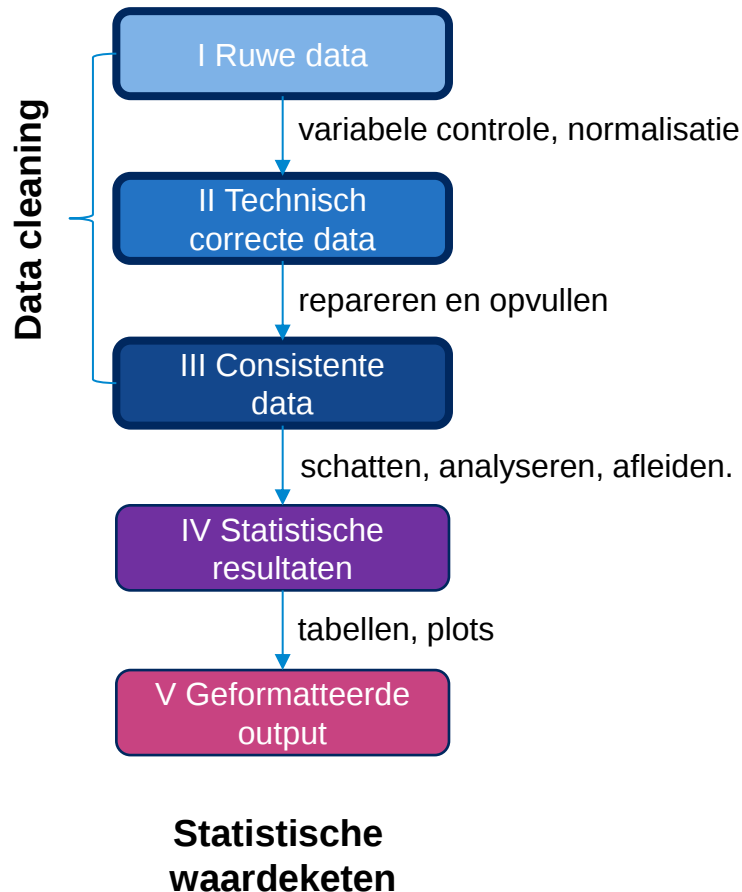
Een imitatie van het 1CHO bestand voor gebruik in het najaar

Dashboards

Op basis van subsets van de analysesets

Extracten voor wetenschappers

Op basis van subsets van de analysesets



Data cleaning is het proces om van ruwe data naar consistente data te komen die geanalyseerd kan worden.

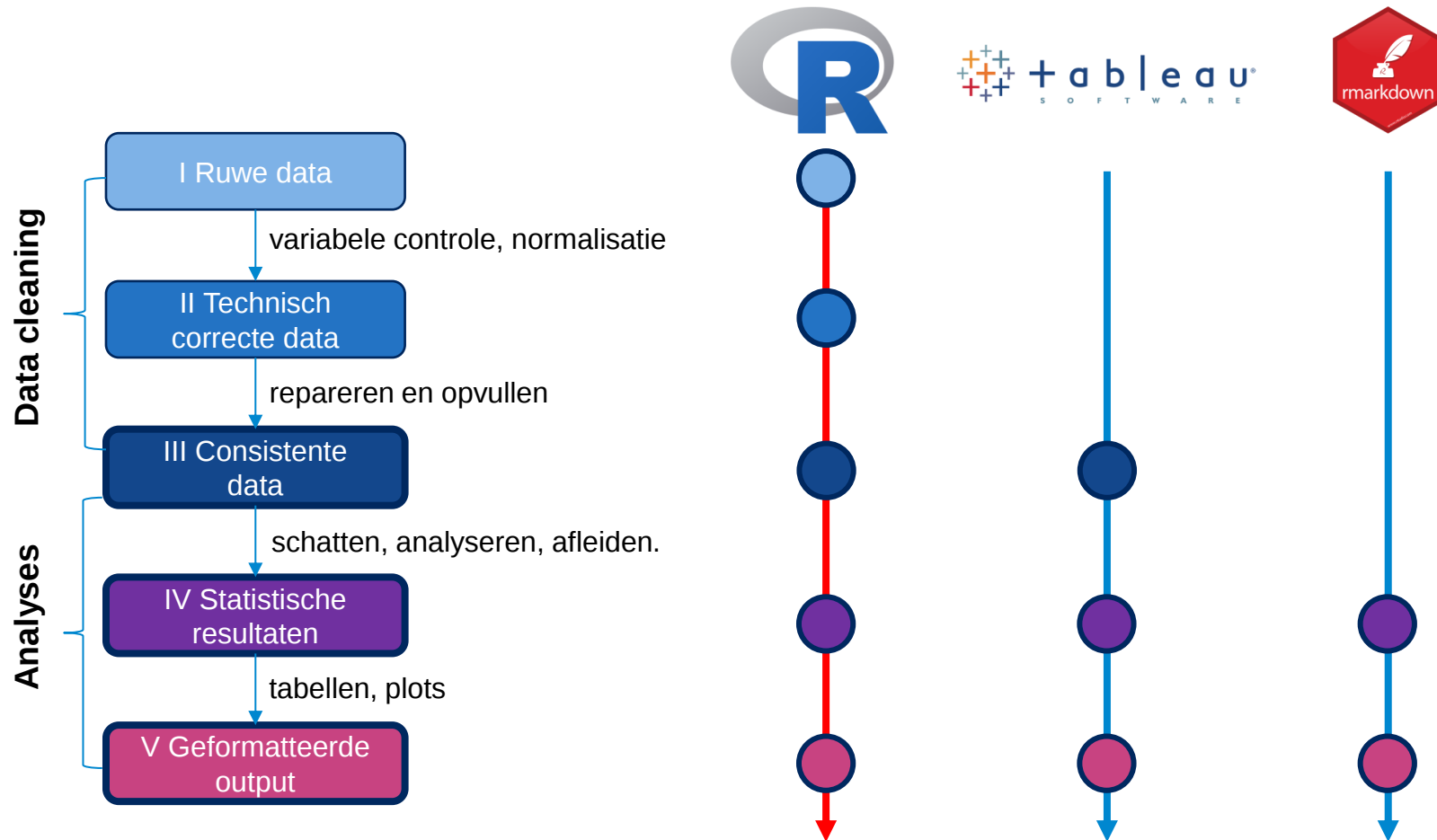
Data cleaning is ook het definiëren van outliers en het opvullen van missende waarden (*imputation*). Het is van invloed op de statistische uitkomsten!

Herhaalbaarheid van data cleaning is daarom een voorwaarde voor herhaalbaarheid van het statistische onderzoek.

Bron: CBS 2013, p7



R & R-studio



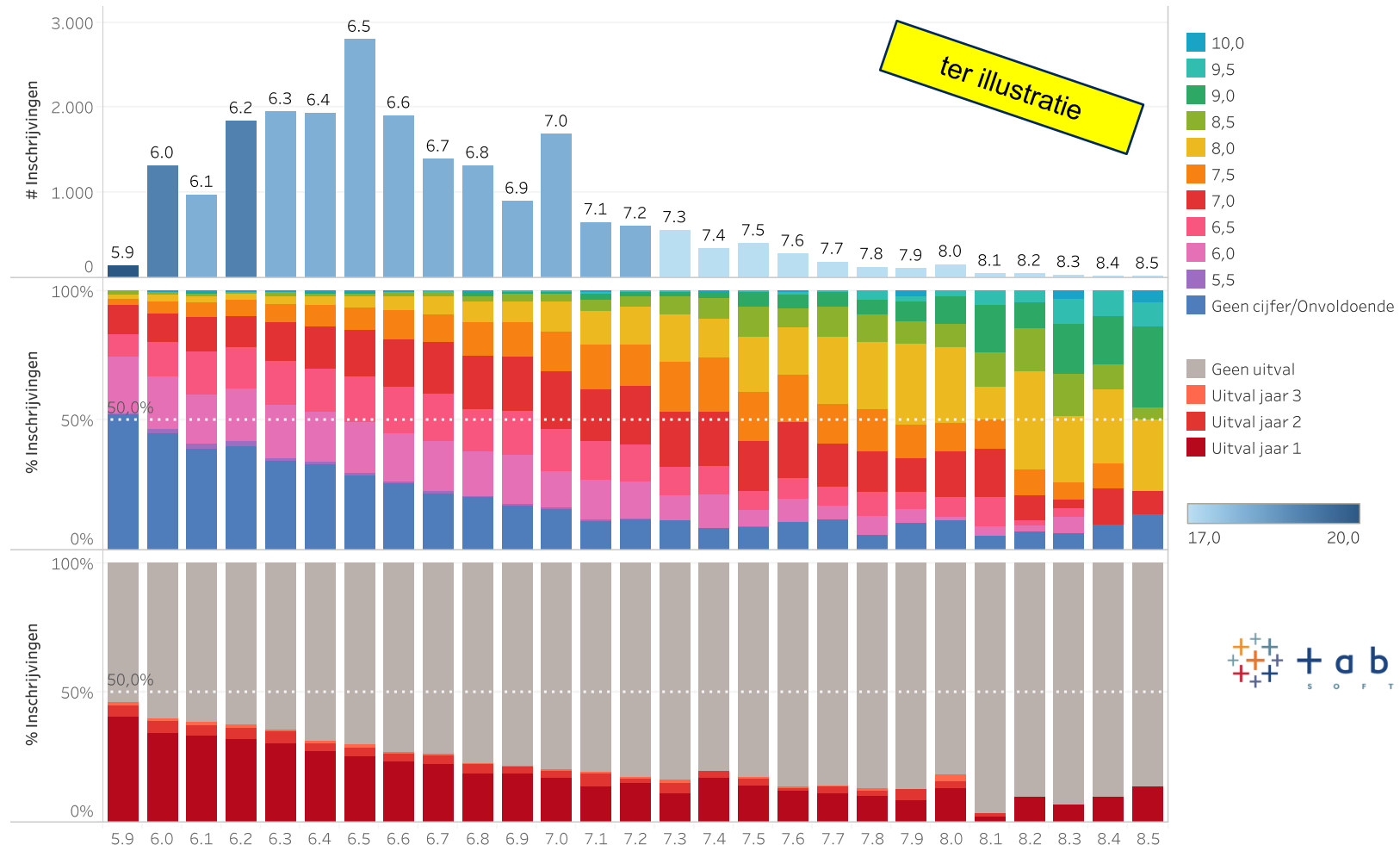
**Statistische
waardeketen**



Met een combinatie aan software analyseren we data en publiceren er onze resultaten mee:
R & R-studio, Tableau Desktop, Markdown.



Tableau gebruiken we voor het snel visualiseren van mogelijke verbanden en toelichtingen in rapporten.





Markdown gebruiken we voor kennisdeling, verantwoording en uitvoerige documentatie van de analyses.

Team

- Kennisdeling in methode en de code
- Logboek van de analyse

Wetenschap

- Geïnteresseerd in de methode
- Verantwoording van onderzoek
- In beleidsrapporten ontbreekt vaak de statische onderbouwing

Beleid

- Onderbouwing moet kloppen om onzekerheid weg te nemen
- Geïnteresseerd in de uitkomsten en toepassing voor beleid

R-Markdown

- Combineert tekst en code



3.1 Selectie van de data

Voor deze analyse is de analyseset uit het project Student Analytics gebruikt. Per combinatie opleiding - student - cohort is er één rij in de analyseset. Dit betekent dat een student die bij meerdere opleidingen is ingeschreven, meerdere keren voorkomt in de analyseset. Om dit te voorkomen kiezen we alleen inschrijvingen uit inschrijvingsjaar 2016

De volgende selecties zijn uitgevoerd:

- Studenten die de NSE hebben ingevuld,
- Inschrijvingsjaar = 2016 (dit is de laatste NSE, NSE2017)
- Bachelor
- Max. studiejaar = nominaal + 1
- Vooropleiding = vwo of hbo-p

HIDE

```
## Laad de analyseset
Analysis_set_1 <- vusa::Get_analyseset(1, print = F)
# Maak een subset van de data, voeg variabele INS_Studiejaar_max toe,
data_subset <- Analysis_set_1 %>%
  filter(NSE_Ingevuld == 1,
         INS_Inschrijvingsjaar %in% c(2015, 2016), # 2015 is de testset
         INS_Studiejaar <= 4,
         INS_Opleidingsfase_BPM == "B",
         INS_Hoogste_vooropleiding_soort_cat %in% c("HBO-P", "VWO"))
```



Afspraken over scripts en versies



R scripts

- Stijlgids met uitwerking van alle stappen
- Altijd uitgaan van de bron
- Manipulaties goed beschrijven
- Onderlinge review van werk

Atlassian



Versiebeheer

- Tegelijk werken in dezelfde scripts
- Wijzigingen bijhouden en eventueel terugdraaien
- Ingrijpende wijzigingen in een aparte branch

```
72 - #####
73 ## 2. BEWERKEN
74 - #####
75
76 - ## =====
77 ##
78 uitwisseling <- uitwisseling %>%
79   mutate(
80     ## SAP-codering bij kolom UIT_Semester:
81     ## 025 = student gaat in het 1e VU-semester op uitwisseling
82     ## 035 = student gaat in het 2e VU-semester op uitwisseling
83     UIT_Semester = case_when(
84       UIT_Semester %in% c("25", "025") ~ "Semester 1"
85       UIT_Semester %in% c("35", "035") ~ "Semester 2"
86     ),
87     grepl("+", UIT_Semester) ~ "Semester 1"
88   )
89   ## Academisch jaar bewerken van format "2014-2015" naar 2014
90   UIT_Academisch_jaar_uitwisseling = as.numeric(substr(UIT_Academisch_jaar,
91     1, 4))
92   ## Op uitwisseling gegaan TRUE of FALSE
93   UIT_Uitwisseling = case_when(
94     grepl("cancelled", tolower(UIT_Geselecteerd_voor_partner)) ~ TRUE
95     grepl("withdrawn", tolower(UIT_Geselecteerd_voor_partner)) ~ TRUE
96     grepl("not selected", tolower(UIT_Geselecteerd_voor_partner)) ~ FALSE
97     TRUE ~ "uitwisseling")
98   ## Verwijder de waarden in de kolom geselecteerd voor partner, die geen
99   ## universiteit betreffen
100   UIT_Geselecteerd_voor_partner = case_when(
101     UIT_Uitwisseling == "uitwisseling" ~ NULL
102     TRUE ~ UIT_Geselecteerd_voor_partner
103   )
104 }
```

Inlezen inschrijvingen 1OKT en 1CHO zijn nu klaar. Test-Analyseset is gemaakt	11 okt 2017 13:18	Je
o.a. VIFs	10 okt 2017 17:13	VI
Toevoegen nieuwe functies aan package om missende waarden op te vullen. Inschrijvingen 1C	10 okt 2017 17:07	Je
WIP Code blokken geshuffeld, code verfijnd	9 okt 2017 18:24	Je
Merge branch 'ontwikkel' of https://bitbucket.org/vustudentanalytics/sa-scripts into ontwikke	9 okt 2017 17:18	VI
kleine aanpassing comments en opleidingsnaam_fix	9 okt 2017 17:17	VI
package functie toegevoegd: fill_missing() Inlezen inschrijvingen herzien.R	9 okt 2017 17:16	Je

```
Parents: 0cb5402719
Author: Jelger van Zaane <j.d.van.zaane@vu.nl>
Date: woensdag 11 oktober 2017 13:18:56
Committer: Jelger van Zaane

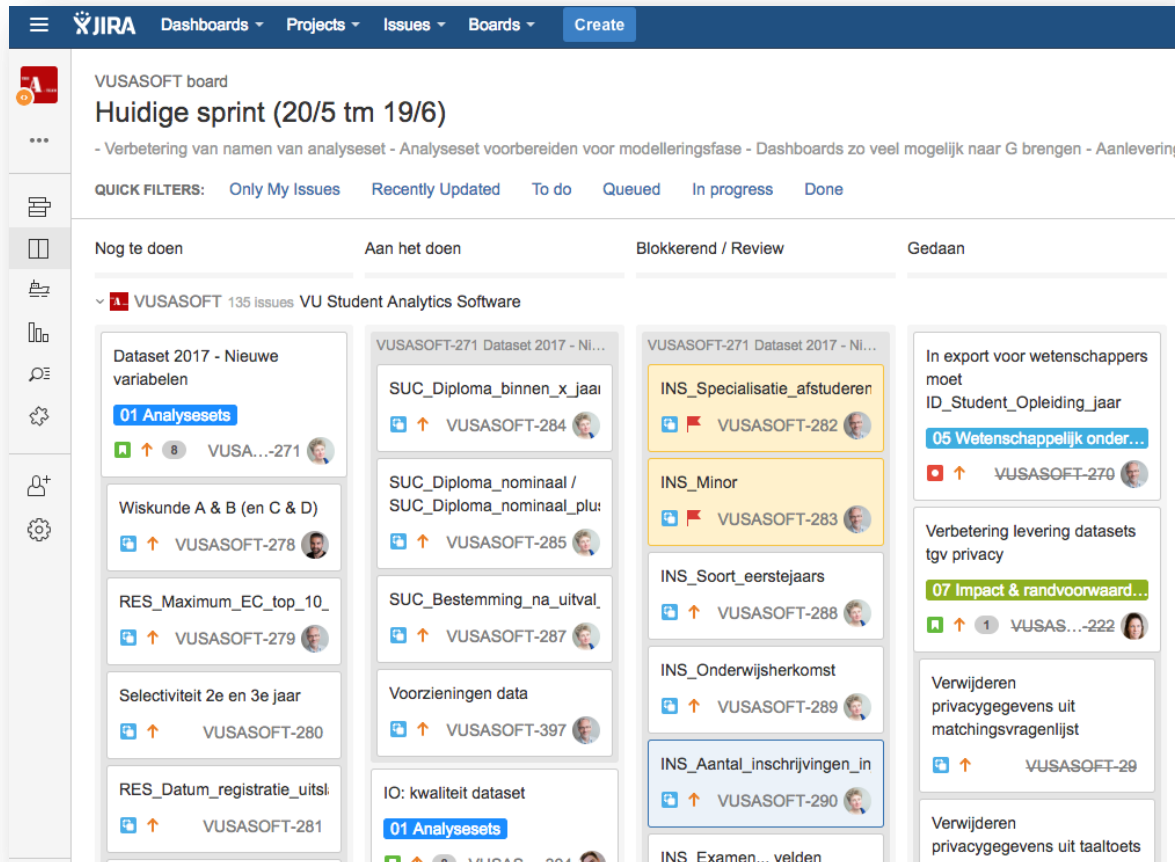
Inlezen inschrijvingen 1OKT en 1CHO zijn nu klaar. Test-
Analyseset is gemaakt

Main script.R
01. Inlezen/Inlezen Inschrijvingen 1CHO.R

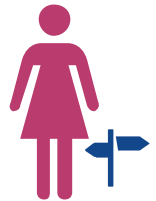
255 255 ## bepaal de icho variabele en controleer of die bestaat
256 256 var_1okt <- paste0(var_1cho, "1OKT")
257 257
258 258 - ## Als de variabele niet bestaat, wordt deze niet opgevuld
259 259 + ## Als de variabele niet bestaat, wordt deze niet opgevuld, maar de 1OKT variant
260 260 + ## wordt wel verwijderd
261 261 + if (is.null(var_1okt)) {
262 262 +   print(paste0(var_1cho, "is niet vervangen omdat er geen icho variabele bestaat"))
263 263 +   df[,var_1okt] <- NULL
264 264 +   print(paste0(var_1cho, "is verwijderd omdat er geen icho variabele bestaat"))
265 265 +   return(df)
266 266 } else{
```



Afspraken over taakverdeling en voortgang



Data Engineer



Business Intelligence Analyst



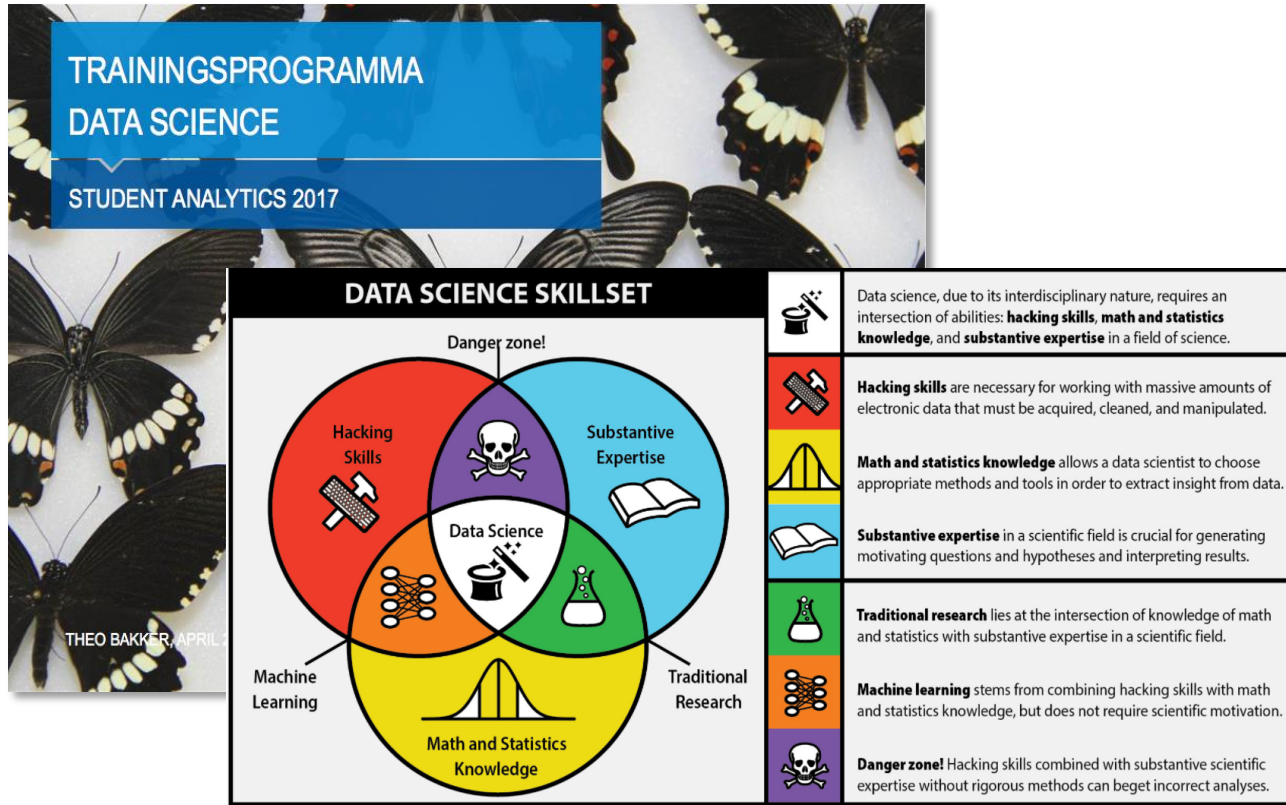
Data Scientist

SCRUM & JIRA

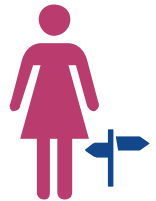
Dagelijkse stand-up / sprints / JIRA



Afspraken over ontwikkeling van capaciteiten



Data Engineer



Business Intelligence Analyst



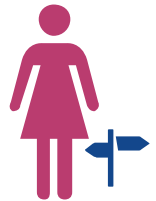
Data Scientist

Training Data Science

Datacamp abonnement, inkoop training Highland Statistics, literatuur, wekelijkse verdiepingssessie, Tableau training, informatiesessie privacy, onderlinge review van werk, wetenschappelijke personeel VU



Data Engineer



Business Intelligence Analyst



Data Scientist

Stijlgids

Afspraak over codering van bestanden en het maken van code



00. Documentatie Analyseset

Start Invoegen Pagina-indeling Formules Gegevens Controleren Beeld

Calibri (Hoofd... 12 A A

Plakken B I U

Tekstterugloop Standaard

Samenvoegen en centreren % 000

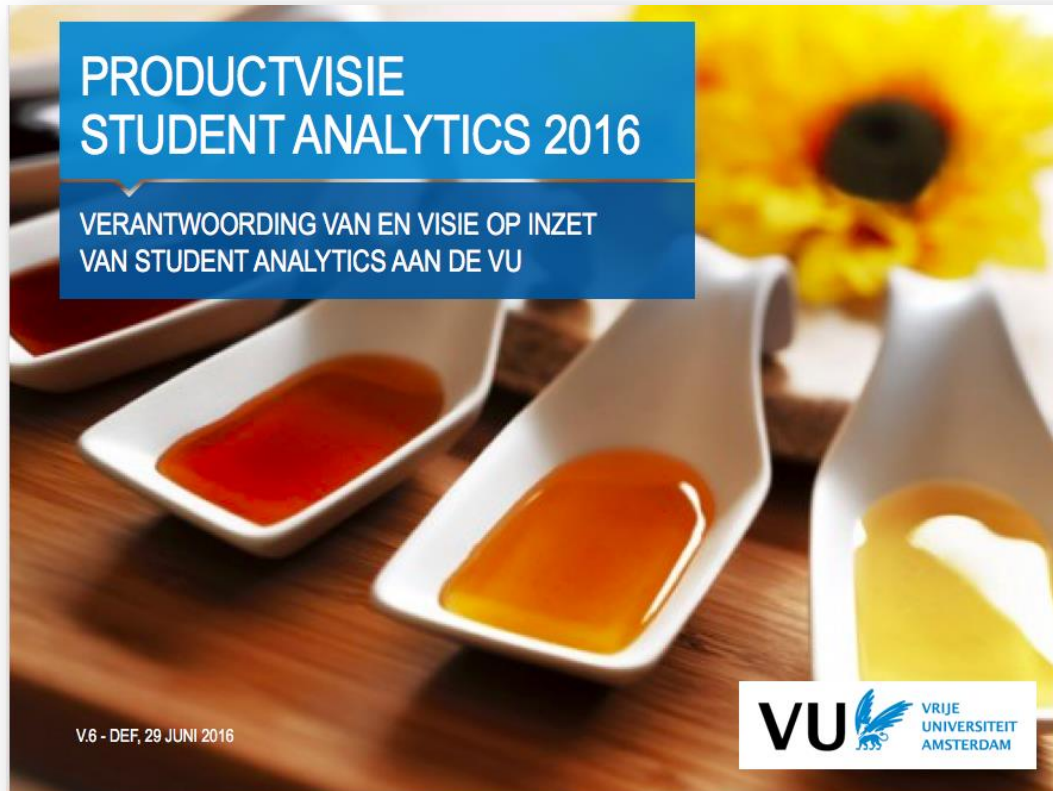
C4 x ✓ fx Omschrijving van de kwalitatieve eis als die er is

	A	B
100	Inschrijvingen	INS_Eerste_jaar_opleiding_en_instelling
101	Inschrijvingen	INS_Eerste_jaar_soort_hoger_onderwijs
102	Inschrijvingen	INS_Examen_na_1_jaar
103	Inschrijvingen	INS_Examen_na_2_jaar
104	Inschrijvingen	INS_Examen_na_3_jaar
105	Inschrijvingen	INS_Examen_na_4_jaar
106	Inschrijvingen	INS_Examen_na_5_jaar
107	Inschrijvingen	INS_Examen_na_6_jaar
108	Inschrijvingen	INS_Examen_na_7_jaar
109	Inschrijvingen	INS_Examen_na_8_jaar
110	Inschrijvingen	INS_Examen_na_9_jaar
111	Inschrijvingen	INS_Examenresultaat
112	Inschrijvingen	INS_Faculteit
113	Inschrijvingen	INS_Geslacht
114	Inschrijvingen	INS_Herinschrijving_na_1_jaar
115	Inschrijvingen	INS_Herinschrijving_na_2_jaar
116	Inschrijvingen	INS_Herinschrijving_na_3_jaar
117	Inschrijvingen	INS_Herinschrijving_na_4_jaar
118	Inschrijvingen	INS_Herinschrijving_na_5_jaar
119	Inschrijvingen	INS_Herinschrijving_na_6_jaar
120	Inschrijvingen	INS_Hoofdneven
121	Inschrijvingen	INS_Hoogste_vooropleiding_binnen_HO_code
122	Inschrijvingen	INS_Hoogste_vooropleiding_BRIN_1CHO
123	Inschrijvingen	INS_Hoogste_vooropleiding_BRIN_SAP
124	Inschrijvingen	INS_Hoogste_vooropleiding_code_1CHO
125	Inschrijvingen	INS_Hoogste_vooropleiding_geverifieerd_SAP
126	Inschrijvingen	INS_Hoogste_vooropleiding_jaar_1CHO
127	Inschrijvingen	INS_Hoogste_vooropleiding_jaar_SAP

Documentatie analyseset (codebook)
Beschrijving per variabele & mogelijke waarden



**Data
Scientist**



Inhoudsopgave

- I. Inleiding
- II. Context
- III. Scope
- IV. Ontwerpcriteria**
- V. Verantwoording keuzes**
- VI. Procedures levering datasets

Productvisie

Verantwoording van selecties en scope van de data



**Business
User**



**Project
Sponsor**



Martijn Meeter
Hoogleraar
Onderwijskunde VU



Sandjai Bhulai
Hoogleraar
Data Analytics VU

Wetenschappelijke toetsing

Methodiek en uitkomsten worden gevalideerd door wetenschappers van de VU



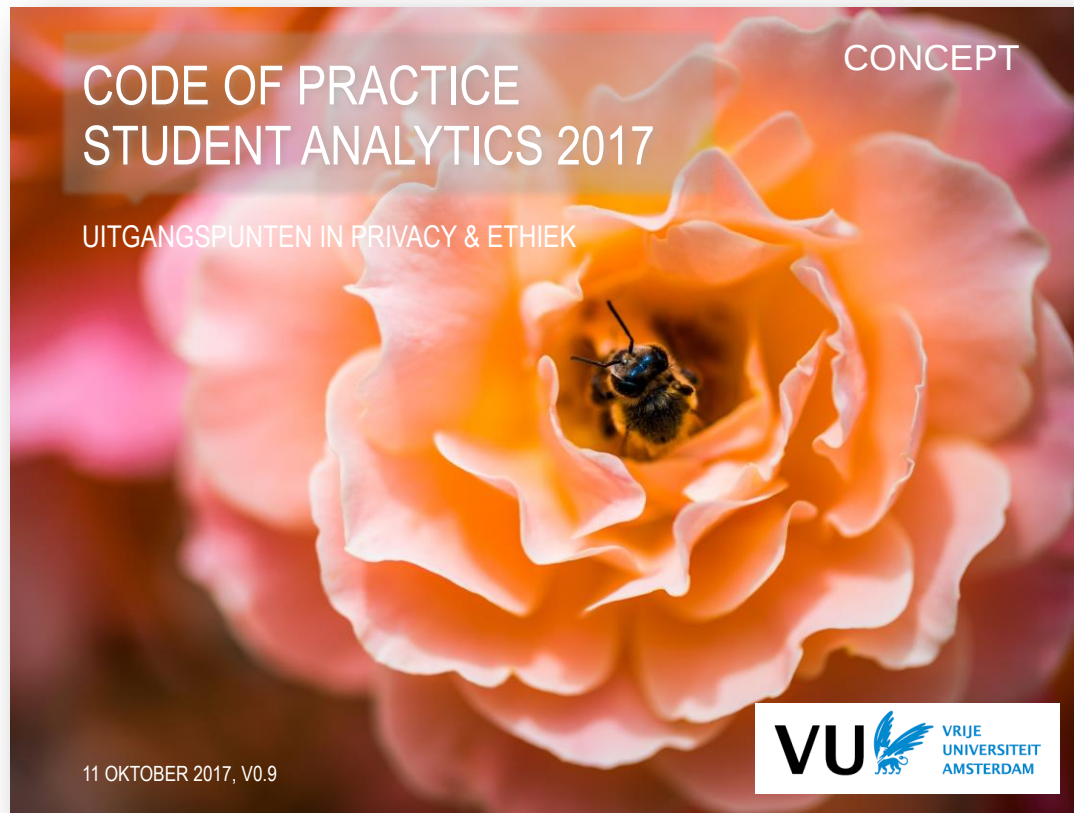
Wetenschappelijke bijlage
voorspelmodel



**Data
Scientist**



Afspraken over wat we wel en wat we niet willen doen

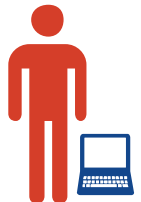


Code of conduct

Waar staat de VU voor in het gebruik van Student Analytics



**Privacy
Officer**



**Business
User**



**Project
Sponsor**



Janneke Lommertzen

Beleidsmedewerker Institutional Research
Student- en Onderwijszaken, Vrije Universiteit
j.lommertzen@vu.nl



Jelger van Zaane

Data engineer Student Analytics / Institutional Research
Student- en Onderwijszaken, Vrije Universiteit
j.d.van.zaane@vu.nl

VOORWAARDEN VOOR GEBRUIK VAN DEZE PUBLICATIE (1/2)



Deze uitgave deelt de VU met externen onder de Creative Commons licentie:
Naamsvermelding-NietCommercieel-GelijkDelen.



Dit is de vereenvoudigde (human-readable) versie van de volledige [licentie](#) en geen vervanging van de volledige licentie. [Vrijwaring](#).

Je bent vrij om:

- **het werk te delen** — te kopiëren, te verspreiden en door te geven via elk medium of bestandsformaat
- **het werk te bewerken** — te remixen, te veranderen en afgeleide werken te maken
- De licentiegever kan deze toestemming niet intrekken zolang aan de licentievoorwaarden voldaan wordt.

Onder de volgende voorwaarden:

- **Naamsvermelding** — De gebruiker dient de maker van het werk te [vermelden](#), een link naar de licentie te plaatsen en [aan te geven of het werk veranderd is](#). Je mag dat op redelijke wijze doen, maar niet zodanig dat de indruk gewekt wordt dat de licentiegever instemt met je werk of je gebruik van het werk.
- **NietCommercieel** — Je mag het werk niet gebruiken voor [commerciële doeleinden](#).
- **GelijkDelen** — Als je het werk hebt geremixt, veranderd, of op het werk hebt voortgebouwd, moet je het veranderde materiaal verspreiden onder [dezelfde licentie](#) als het originele werk.
- **Geen aanvullende restricties** — Je mag geen juridische voorwaarden of [technologische voorzieningen](#) toepassen die anderen er juridisch in beperken om iets te doen wat de licentie toestaat.

VOORWAARDEN VOOR GEBRUIK VAN DEZE PUBLICATIE (2/2)



(Vervolg)



Let op:

Voor elementen van het materiaal die zich in het publieke domein bevinden, en voor vormen van gebruik die worden toegestaan via een [uitzondering of beperking](#) in de Auteurswet, hoeft je niet aan de voorwaarden van de licentie te houden.

Er worden geen garanties afgegeven. Het is mogelijk dat de licentie je niet alle gebruiksvrijheden geeft die nodig zijn voor het beoogde gebruik. Bijvoorbeeld, andere rechten zoals [publiciteits-, privacy- en morele rechten](#) kunnen het gebruik van een werk beperken.

De volledige versie van de licentie op deze publicatie is van toepassing. Zie <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.nl>