

EEN COMBINATIE VAN MULTIPLE IMPUTATION (MI) EN LATENTEKLASSENANALYSE (LC) OM TE CORRIGEREN VOOR MEETFOUT

LAURA BOESCHOTEN (UVT & CBS)

DANIEL OBERSKI (UU)

TON DE WAAL (CBS & UVT)

JEROEN VERMUNT (UVT)

VOLKSTELLING

2001: Eerste 'virtuele' volkstelling

- Data is al beschikbaar in registers
- Gebruik hiervan is veel goedkoper dan het uitvoeren van een volkstelling
- Weerstand onder de bevolking (met unit-nonresponse tot gevolg)



VOLKSTELLING

Eisen leiden tot methodologische problemen

- Een groot aantal grote tabellen moet geproduceerd worden
- Resultaten in deze tabellen moeten numeriek consistent zijn
- Er moet rekening worden gehouden met onmogelijke combinaties van scores

VOLKSTELLING

Combineer data uit meerdere bronnen

- GBA (Gemeentelijke BasisAdministratie) ← populatie
- EBB (Enquete BeroepsBevolking) ← steekproef
Personen ontbreken in de EBB

Wegen leidt tot inconsistenties over de verschillende tabellen

Gevolg: Zowel vooraf als achteraf wordt er een hoop aan de data gesleuteld



Een efficiënte procedure is hier nodig

DAARBIJ IS EEN METHODE NODIG DIE ...

1. ...zorgt voor consistente schattingen
2. ...edit restricties automatisch meeneemt
3. ...corrigeert voor meetfouten in de data
4. ...onzekerheid door missende en foute waarden meeneemt

OPLOSSING: EEN COMBINATIE VAN...

Latenteklassenanalyse

1. Gebruik variabelen die hetzelfde meten als indicatoren van een latente variabele die de 'echte waarden' meet
2. Specificeer edit restricties om onmogelijke combinaties tussen de latente variabele en andere variabelen te voorkomen.

Multiple imputation

1. Zorgt voor consistente schattingen omdat we de hele populatie imputeren en niet meer hoeven te wegen
2. Onzekerheid door missende en conflicterende waarden wordt meegenomen

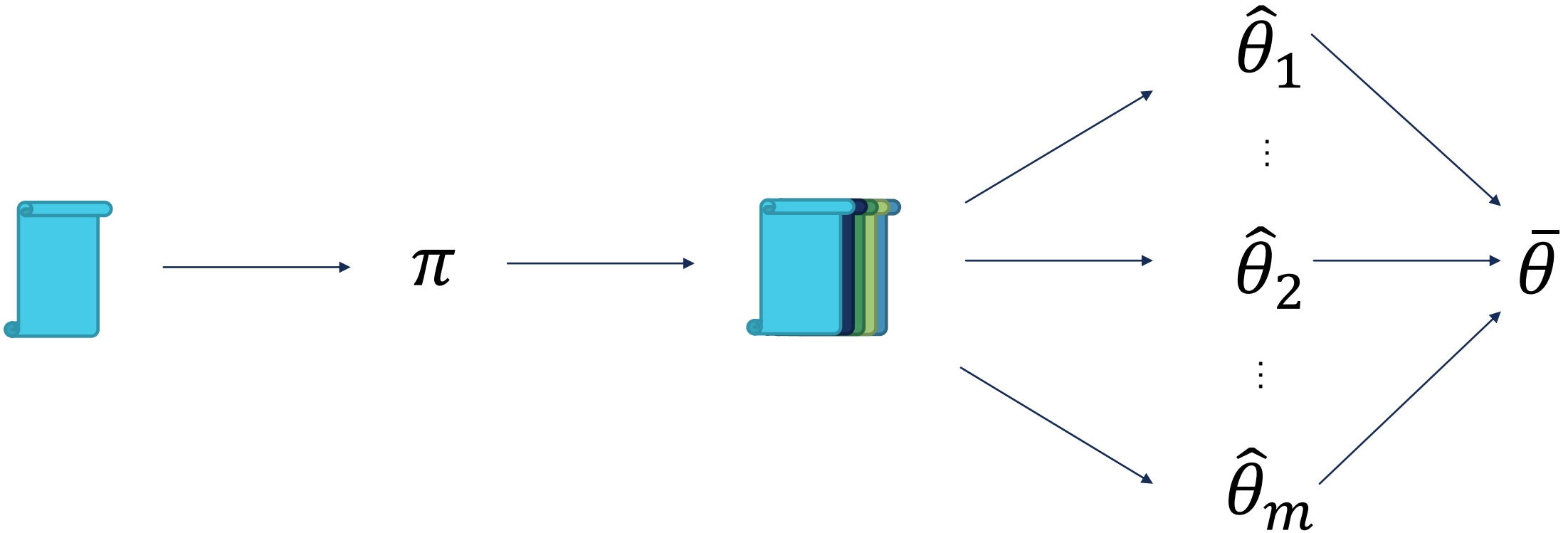
OVERZICHT VAN DE MILC METHODE

1. LC model

2. Imputaties

3. Schattingen

4. Pool



CONCLUSIES

- Kwaliteit van de geïmputeerde variabele hangt samen met de kwaliteit van de data
- Dit wordt samengevat in de entropy R^2
- Neem alle relevante variabelen mee in het LC model als covariaat
- Nog niet duidelijk hoe de methode met lastigere situaties omgaat

SWOV

STICHTING WETENSCHAPPELIJK ONDERZOEK VERKEER

Doel

Schat het aantal gewonden in verkeersongevallen per vervoerstype in 1994

Abbreviated injury Score

AIS	Injury	Example	% death
1	Minor	superficial laceration	0
2	Moderate	fractured sternum	1–2
3	Serious	open fracture of humerus	8–10
4	Severe	perforated trachea	5–50
5	Critical	ruptured liver with tissue loss	5-50
6	Maximum	total severance of aorta	100

DATA

Ziekenhuis

- Vervoerstype
- Provincie



- Gelinkt
- Incompleet
- N=16.962

We hebben ook:

- Leeftijd
- Geslacht
- Externe info
- AIS-score

Politie

- Vervoerstype
- Provincie

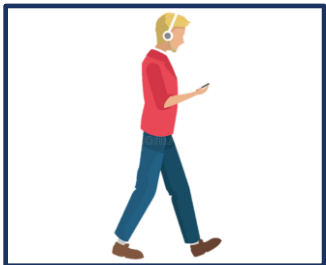


CATEGORIEËN

- Ongelukken met gemotoriseerde voertuigen



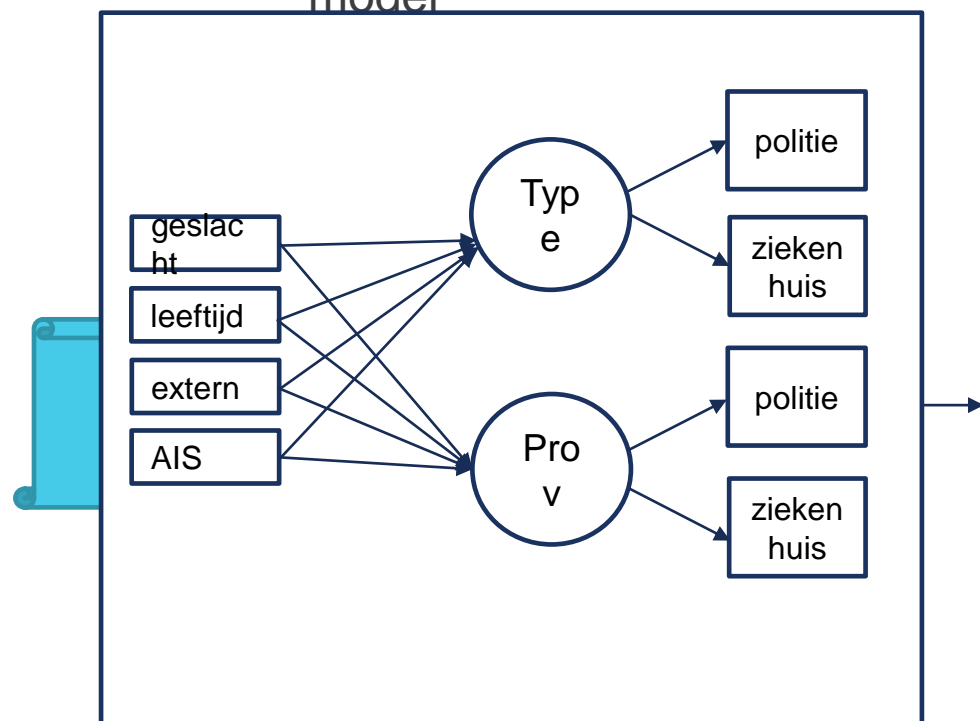
- Ongelukken zonder gemotoriseerde voertuigen



7 klassen

TOEPASSEN VAN DE MILC METHODE

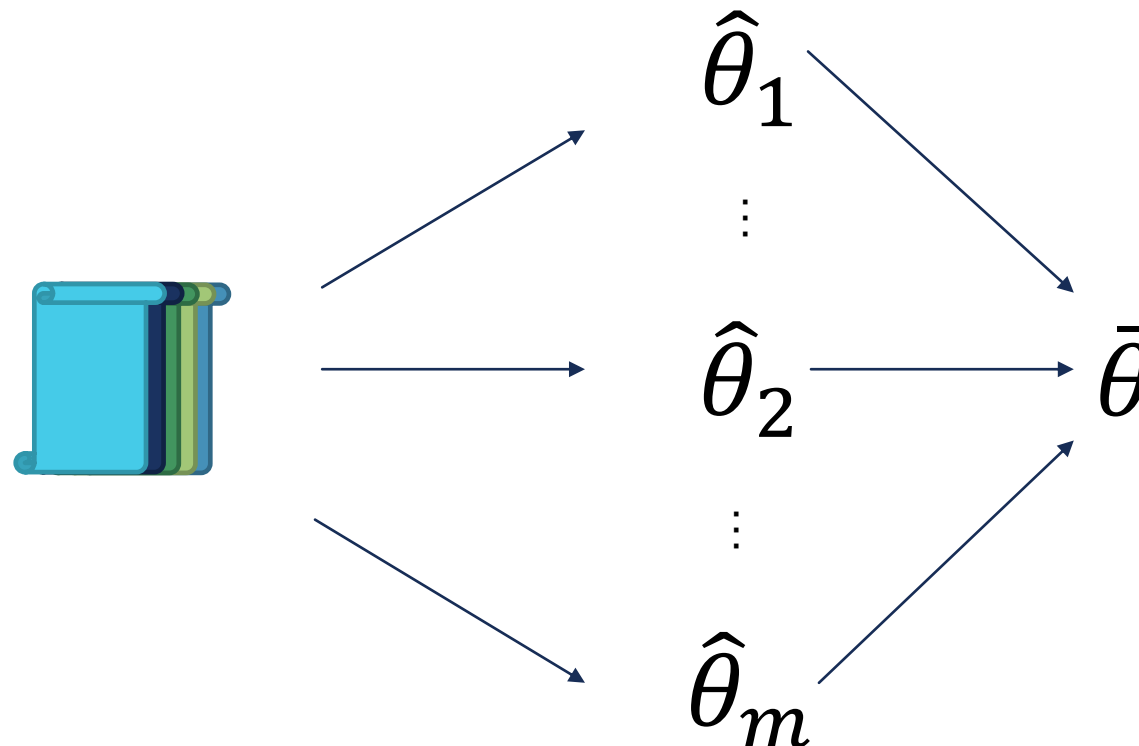
1. LC
model



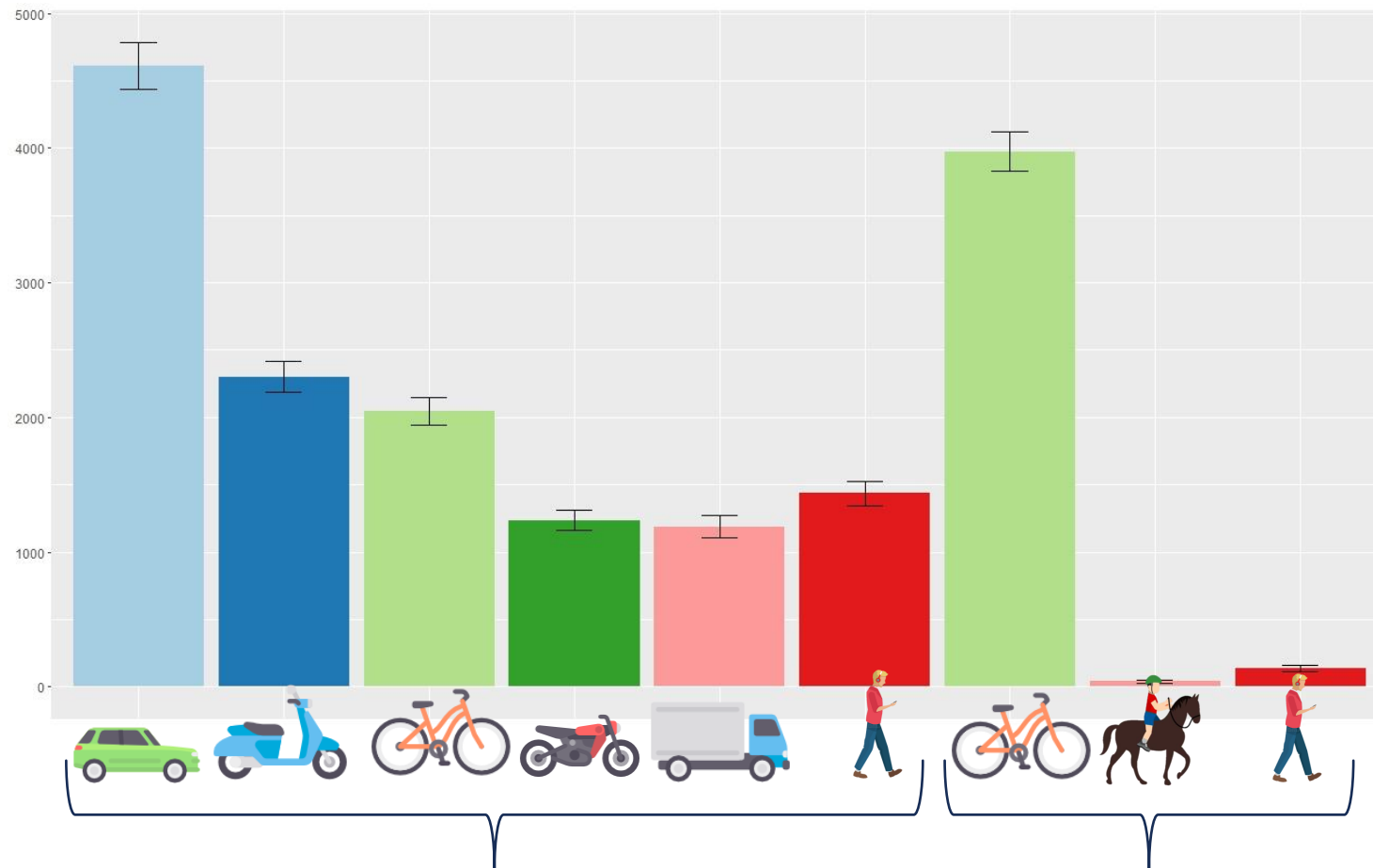
2. Imputaties

3. Schattingen

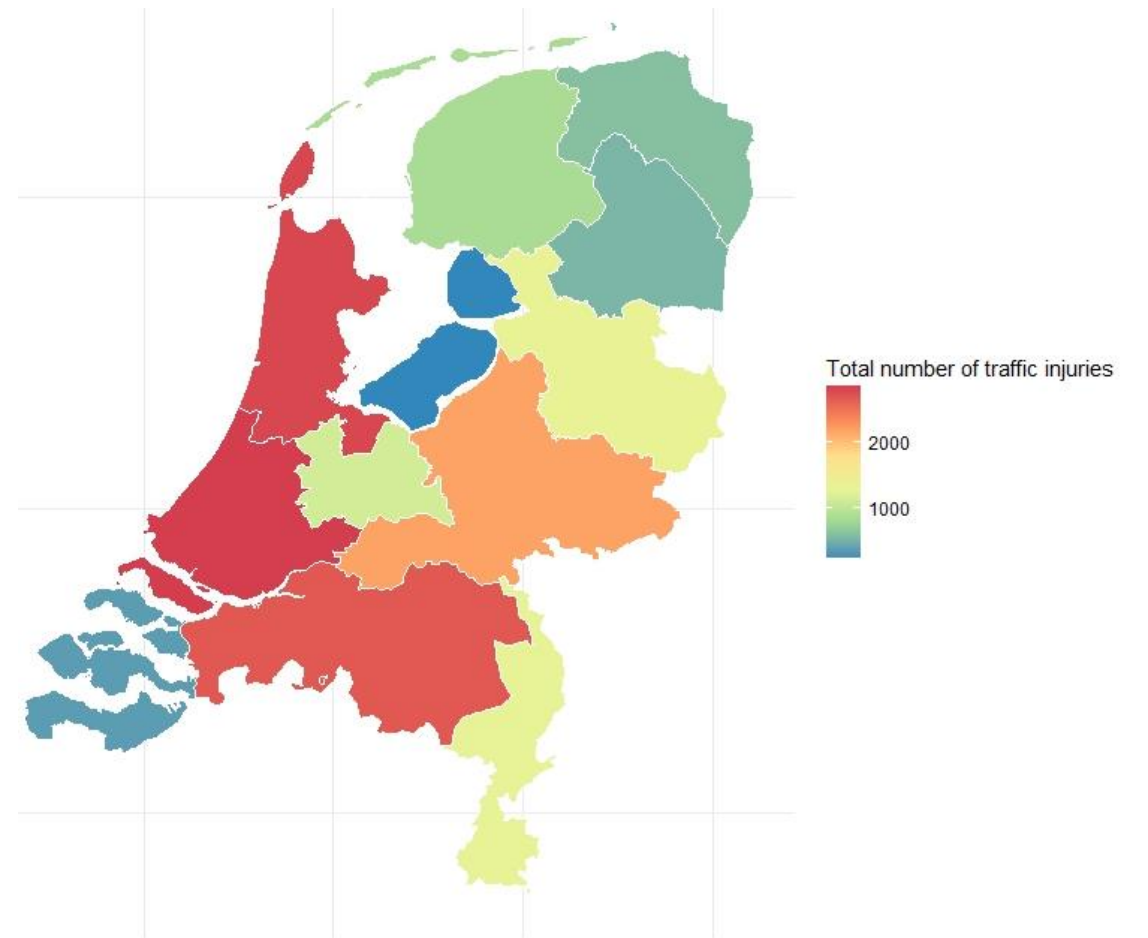
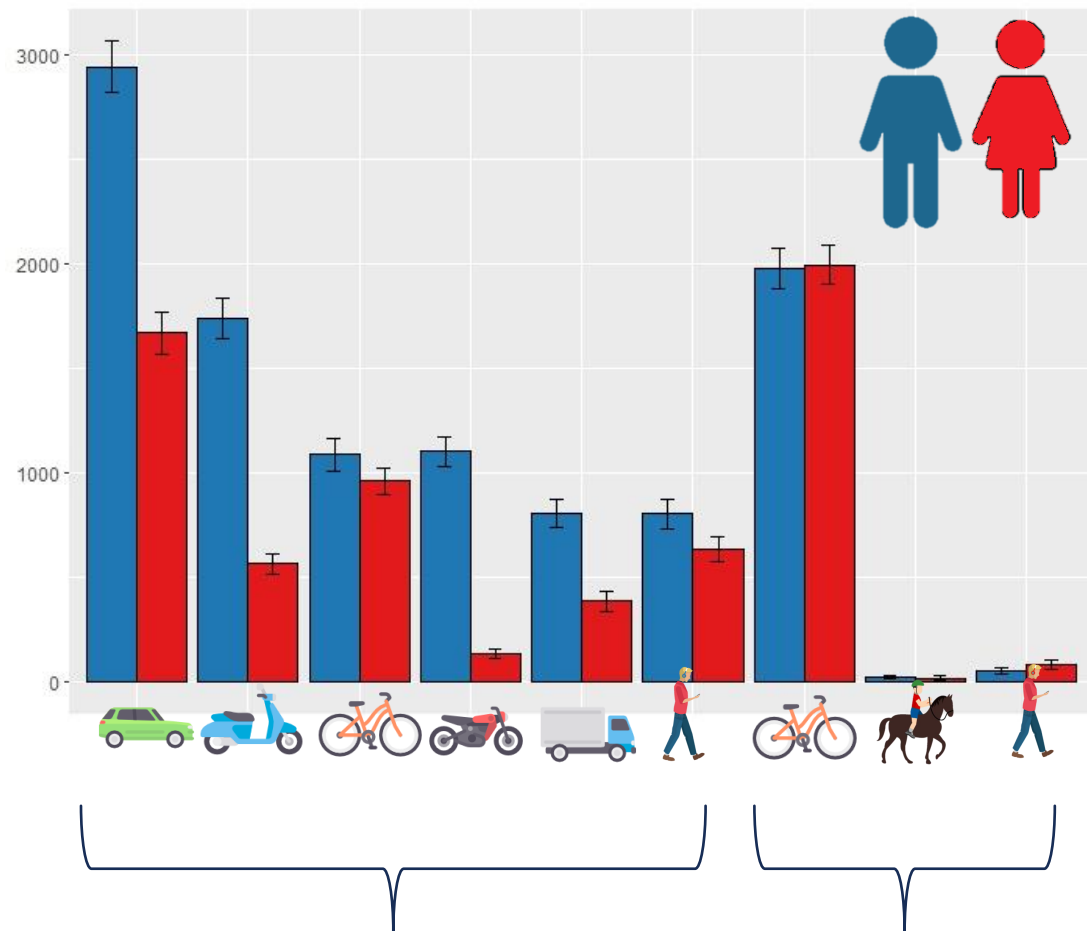
4. Pool



OUTPUT

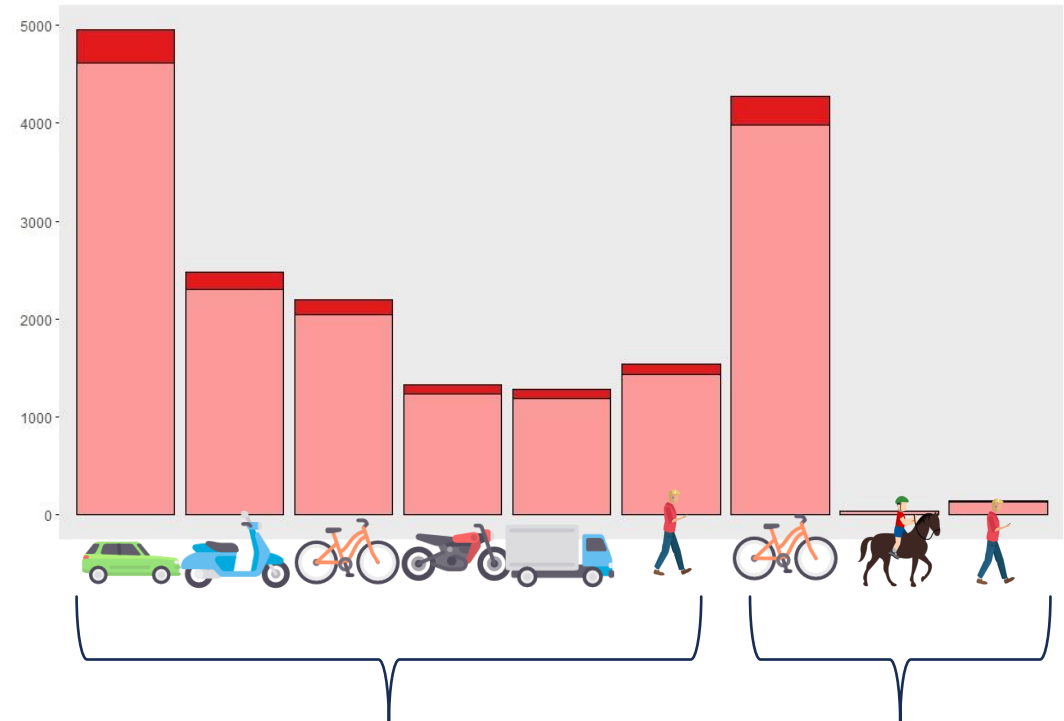


OUTPUT



IDEE: MILC MET C/RC COMBINEREN

- Schat het aantal verkeersgewonden niet door politie of ziekenhuis geregistreerd
- Gebruik LC model om categorieën toe te wijzen aan de bijgeschatte cases
- Gebruik bootstrap om de onzekerheid te schatten



HOE GAAT MILC OM MET LASTIGE SITUATIES?

Zoals ...

- # observaties
- # variabelen
- # categorieën
- # edit restricties
- # latente variabelen
- Unit missingness in surveys
- Unit missingness in registers

SWOV

Toepassing



CENSU

S

Simulatie



DANK!