|  |
| --- |
| Course: Data Science [Z15811] Class: Setting the Scene [01]  Subject: Discussion topics |

1. ***Artificial intelligence, machine learning, deep learning and data science are buzzwords that people love to use interchangeably, but they are not entirely the same. Do some research and write a definition that makes the differences clear.***

AI: Lost problemen op met recursieve problemen.

^

Machine learning: Leren zelfstandig doen

^

Deep learning: Leren zelfstandig denken

Data science: Data analyseren

1. ***In this buzzword bingo, what is the place of data engineering? What problem does it solve?***

Data engineers are responsible for finding trends in data sets and developing algorithms. Structures the data to 1 big data pool.

1. ***In structured, tabular data (also called rectangular data), you can have different sorts of data. Come up with 3 different types.***

* Integer
* Float
* String

1. ***Explain the difference between numerical and categorical data. Make up 3 examples of structured datasets that contain both numerical and categorical data.***

Numerical kan genummerd en geor worden (lengte, score, volume,…)  
Categorical niet (kleur, geslacht, type,…)

* MPG dataset (auto’s)
* Restaurant menu
* Mensen

1. ***Can you use*** ggplot2’s facet\_wrap() ***function on every variable in a dataset, regardless of their type? Why (not)?***

Neen, dit is enkel nuttig met een beperkt domein, teveel en het wordt onduidelijk. Continue variabelen gaan dus niet.

1. ***Experiment with*** ggplot2, ***and map a continuous variable to color, size, and shape. How do these aesthetics behave differently for categorical vs. continuous variables?***

Bij beperkte domeinen (categorical) krijg je duidelijke verschillen.

Met continue variabelen zijn deze verschillen graduaal.

1. ***Find out which smoothing method*** geom\_smooth() ***uses to plot a trend line.***

GLM, Generalized Linear Mix