Week 1:

To do:

* github werkend krijgen
* correct data format
* data cleaning
  + missing values
  + correcting data
* Data transformation
  + normalization
  + aggregation
  + generalization
  + attribute construction

(bijvoorbeeld)

Meeting 1 deliverables:

* classificeren van de data (zie hoorcollege slides)
* cleaned dataset
* lijst met stappen die zijn genomen
* bevindingen tijdens het cleanen van de dataset
* plan van aanpak

Data mistakes:

* incident\_id: lijkt geen problemen in te zitten
* date: deze column is compleet
* state: deze column is compleet
* city\_or\_county: deze column is compleet (bij een aantal staat er wat achter de naam

tussen haakje weet niet of dit een probleem is)

* address: bij sommige staat er een witregel en dus onbekende data
* n\_killed: lijst is volledig
* n\_injured: lijst is volledig
* incident\_url: lijst is compleet al misschien niet erg nuttig
* source\_url: heeft soms een witregel is ook misschien niet zo nuttig
* incident\_url\_fields\_missing: geeft false op iedere rij omdat de incident\_url compleet is
* congressional\_district: geen witregels
* guns\_stolen: bijzonder incompleet
* gun\_type: net zo incompleet als guns\_stolen
* incident\_characteristics: zo nu en dan een witregel

Legenda:

Groen = gedaan

Blauw = Tessel

Geel = Raymon

Roze = Jeroen

Oranje = Chang

Grijs = Iedereen

Woensdag 6 juni:

Beginnen met data cleanen

Donderdag 7 juni:

Dataset volledig gecleaned

Vrijdag 8 juni:

Beginnen data transformation

Maandag 11 juni:

Data transformation afgerond

Voorbereiden meeting