## Příprava dat:

* Načtení
* *V prezentaci lépe popsat csv soubor*
* Missing values
* Nečíselné hodnoty → odstraněno 86 řádků (84 144 → 84 058)
* Vizualizace spekter a jejich mediánů (+ srovnání mediánů v jednom grafu)

## K-nejbližších sousedů

Testovací subjekt je zařazen do třídy, která převažuje mezi jeho k nejbližšími sousedy (nejsou nutné předpoklady o rozložení, méně citlivé na odlehlé hodnoty jako jeden soused, problém při nevyváženém počtu objektů v jednotlivých skupinách)

* Pro všechny je vyčíslená accuracy při pěti sousedech, vykreslený graf jak se měnní accuracy s počtem sousedů
* Všechny proměnné --5n---0,982
* Bez proměnných pro souřadnice --5n---0,987
* Každá 5tá proměnná --5n---0,988
* Každá 10tá proměnná --5n---0,987
* *Accuracy se s méně proměnnými zvyšuje (overfitting)*

## Rozhodovací strom

Postupné rozdělování datasetu podle hodnot jednotlivých proměnných.

* Pro všechny je/bude získaná accuracy a nákres stromečku
* Všechny proměnné --0,962
* Bez proměnných pro souřadnice --0,962
* Každá 5tá proměnná --0,963
* Každá 10tá proměnná --0,886
* *Podle počtu vrstev ještě*

## Centroidová metoda

Objekt zařazen do skupiny s nejbližším centroidem (*popsat co je centroid*). Bez předpokladů, vyřeší problémy sousedů – nevadí nevyvážený počet ve skupinách.