Práce s (kvantitativními) daty

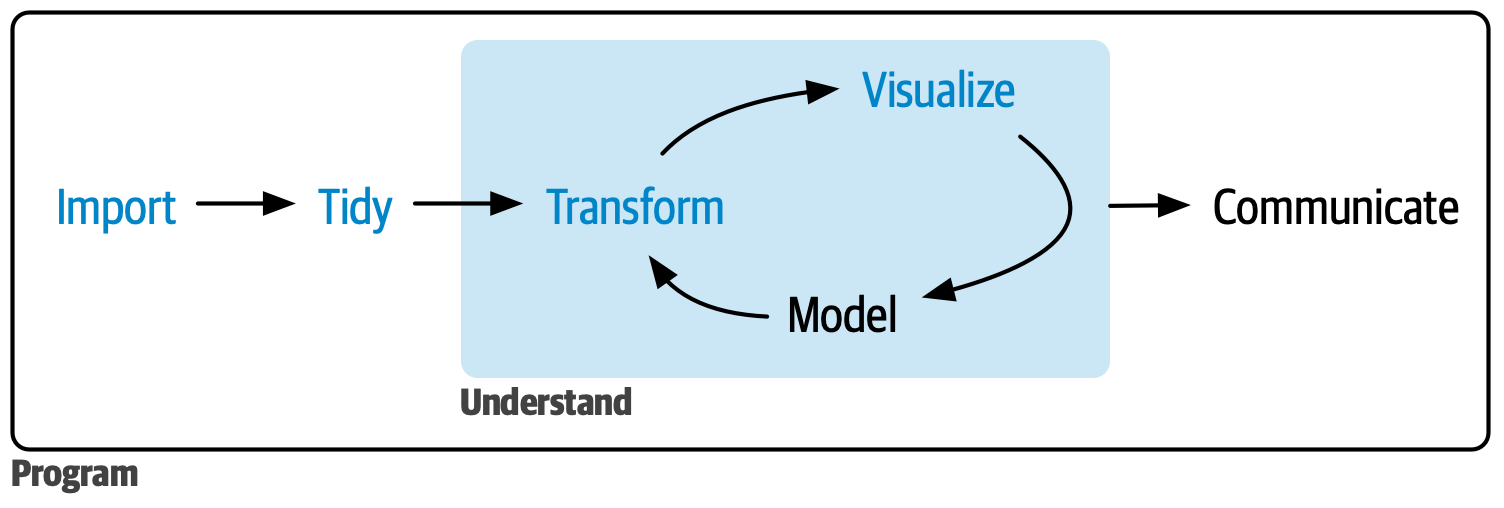
Subtitle

## Jak vypadá datová analýza dnes

* kód, reprodukovatalné [workflow](https://data.gov.cz/datov%C3%A9-sady)
* důraz na uživatele, porozumění jeho potřebám
* iterace, agilní řízení
* produkty, ne projekty
* nejen reporty, mizí zeď mezi analytickým postupem a výstupem

důvěru získáte schopností reagovat na potřeby, transparentností a integritě postupu a komunikaci výsledků

## Mentální model postupu práce s daty



## Projektová a pracovní rovina

* ustálení základních postupů a standardů - pro sebe nebo v týmu - vám umožní soustředit se na důležitá obsahová/analytická rozhodnutí namísto těch triviálních a sníží vám bariéru k tomu tyto praktické principy dodržovat
* mezi ty patří
  + předem daný způsob dokumentace projektu, dat, postupu (readme, krycí list datového souboru)
  + organizace a názvy souborů v projektové složce
  + názvy proměnných
  + umožnit sám sobě snadný způsob, jak okamžitě zachycovat poznámky z postupu práce - cokoli zjistíte a později byste zapomněli, hned někam zapište.

## Datová rovina

* Při hledání dat je vhodné postupovat systematicky a brát v potaz, kde se jaká data berou a jak spolu souvisí data z různých zdrojů
  + např. ČSÚ některá data publikuje jako otevřená data; mnohá data ČSÚ jsou prezentována i v katalogu Eurostatu, často v analyticky přívětivější formě nebo v alternativních agregacích, které se vám mohou hodit.
* Pro analytickou práci je vhodné využívat data v otevřených, standardizovaných formátech: u ČSÚ a Eurostatu otevřená data
* Vhodnou první cestou k datům ČSÚ a Eurostatu jsou jejich katalogy
* Data o životním prostředí z zdravotnictví hledejte u jiných zdrojů (CENIA resp. ÚZIS), jsou spravována jinak než jiná veřejná data
* Číselníky typicky spravuje ČSÚ, používejte ty správné a čerstvé. Najdete je v databázi metainformací. Prostorová data hledejte na ČÚZK - většina už je dostupná v otevřené formě.
* Spolu s daty stahujte (a čtětě!) dokumentaci, abyste rozuměli, jak data vznikla, co v nich je a není, v jakém aspektu se na ně dá a nedáb spolehnout.
* Pokud data sbíráte nebo vytváříte nové datové sady spojováním a transformací jiných, stáváte se de facto správci dat (ne v právním slova smyslu, ale v praktickém). Dokumentujte, jak data vznikla, odkud pocházejí, jak je použít

## Technická rovina

* Využití nástroje PowerQuery vám pomůže
  + uvést do praxe principy moderní práce s daty: postup, dokumentace, transparentnost postupu, reprodukovatelnost a zkontrolovatelnost, automatizace, nedestruktivní přístup ke zdrojovým datům
  + osvojit si postupy užitečné při přechodu k práce s daty v programovacích nástrojích a v databázích
  + provést některé operace, které v Excelu nejdou
  + naladit se na logiku práce s daty v PowerBI
  + využít datové zdroje ve formátech, které z Excelu nejsou dostupné
  + více se soustředit na to, co data říkají, než jak s nimi operovat
* Na pokročilé vizualizace, složitější statistické modelování, ještě větší nebo komplexnější data nebo složitější projekty je vhodné použít jiné nástroje