

# Volební data management

Úvod

# Co je datamangement?

- Činnosti spojené se získáváním, uchováváním a využíváním dat
- Správa dat – strukturování datasetu, pojmenovávání proměnných a dat, tvorba metadat, publikování dat – tomu se budeme věnovat jen okrajově
- Využití dat => analýza – viz další kurzy
- Získávání a propojování dat z různých zdrojů

# Co se v kurzu naučíte?

- Kde a jak najít politologická data
  - Jak stáhnout data z internetových stránek
  - Jak propojovat data z různých zdrojů
  - Jak data „vyčistit“ k analýze
- 
- Zejména s pomocí softwaru R

# Jak se to naučíte

- Praktické semináře
- Průběžné úkoly
- Reflexe průběžných úkolů
  
- Učení se praktickým věcem je kolektivní záležitost
  - Na průběžných úkolech spolupracujte
  - Závěrečný úkol každý sám za sebe

# Průběžné úkoly

- 6 úkolů
- Data odevzdání v sylabu
- Hodnocení splněno/nesplněno/neodevzdáno
- Nesplněné a neodevzdané úkoly nutné přepracovat do 7 dnů
- 6 splněných úkolů = 40 bodů

# Závěrečný úkol

- Dostanete nějaký problém a musíte ho vyřešit
- Když se zaseknete, můžete využít nápovědu
- Jedna nápověda „zdarma“, další za bodovou srážku
- Nápovědy lze získat také za plnění dobrovolných úkolů
- Hodnotící matice bude k dispozici před posledním seminářem
- Max 60 bodů

# Příklad zadání z loňského roku

- Vytvořte tabulku s alespoň 6000 případy, která bude obsahovat informaci tom, jaký dojem měli voliči z kampaní k různým volbám konaných v ČR v letech 2010 – 2018, jakou ve volbách volili (nebo preferovali) stranu/kandidáta. Doplňte informaci o typu voleb a u prezidentských voleb pak typ nominace preferovaného kandidáta (občanská/stranická), jeho procentuální zisk a datum narození.
- Ve vytvořené tabulce ilustrujte libovolnou proměnnou (nebo vztah mezi proměnnými) libovolným grafem.

# Co se v kurzu nenaučíte

- Legální aspekty nakládání s daty pro komerční účely
- Datové „archivnictví“
- Publikování dat
- Analýzu dat



# Čím začít

- Co chceme zjistit?
  - Teoretický nebo „praktický“ problém
  - Teoretický problém – identifikace konceptů a jejich operacionalizace
    - Hledání dat odpovídajících konceptům
  - Praktický problém – možnost vtáhnout k teoriím
    - Možnost induktivního postupu
- Co jsme ochotni podniknout pro získání dat?
  - Hledání existujících dat x vytváření nových
  - Vytváření nových dat je časově, finančně i jinak náročné
- Co musíme s daty provést, abychom mohli zjistit to, co chceme zjistit?

# Jaká data můžeme mít?

- Správná X špatná
  - Závisí na způsobu sběru a vztahu dat a toho co mají měřit
  - Např. odpovědi na otázku Jak moc vás ve volbě strany ovlivnila předvolební kampaň?
- Perfektní data neexistují – nutnost kompromisu
  - Počet případů
  - Vzdálenost mezi konceptem a měřením
  - Čas
  - místo

- Kvalitativní x kvantitativní
  - Kvalitativní data lze obvykle vytvořit s relativně nízkými náklady
    - Rozhovory, focus group, obsahová analýza textu
    - Obtížná a často nejednoznačná interpretace
    - Nemožnost zobecnitelnosti
    - Problematické body: anonymizace, ochrana respondentů
    - Jen zřídka dostupná už vytvořená kvalitativní data
  - Kvantitativní data lze najít pro široké spektrum otázek už vytvořená
    - Vytváření dat je poměrně nákladné
    - 1000 respondentů v online dotazníku stojí okolo 50 - 100 000 Kč
      - Osobní dotazování je v řádech statisíců
    - Vhodné zejména k testování hypotéz nebo popisu populace

- Agregovaná x individuální
  - Agregovaná data jsou k volbám velmi dobře dostupná
  - Individuální data umožňují lépe zjistit, to co chceme zjistit
    - Např. psychologický efekt volebního systému
    - Nebo kdo jsou voliči různých stran
  - Individuální data můžeme agregovat sami (co si myslí veřejnost?)