

Co je statistika, proměnné, úrovně měření

POLb1139 Statistické myšlení v sociálních vědách

To nejdůležitější na začátek?

- Statistika je matematická disciplína

X

- Aplikace statistických metod se bez matematiky do značné míry obejde

Statistika je všude kolem nás

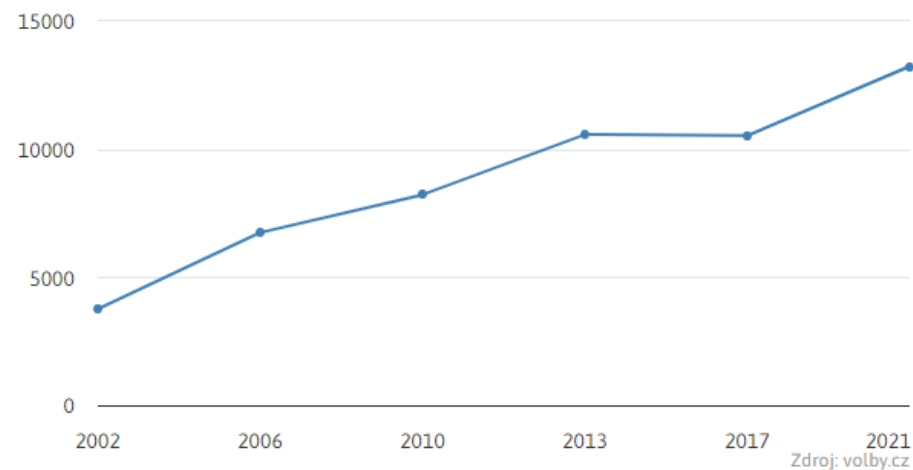
ROZDĚLENÍ EVROPOU

Kam
atříte vy?



Účast ve sněmovních volbách v zahraničí

Hlasů ze světa postupně přibývá, stále ale jde o jednotky tisíc. Z 5,4 milionů hlasů ve sněmovních volbách 2021 představovaly jen 0,2 procenta.

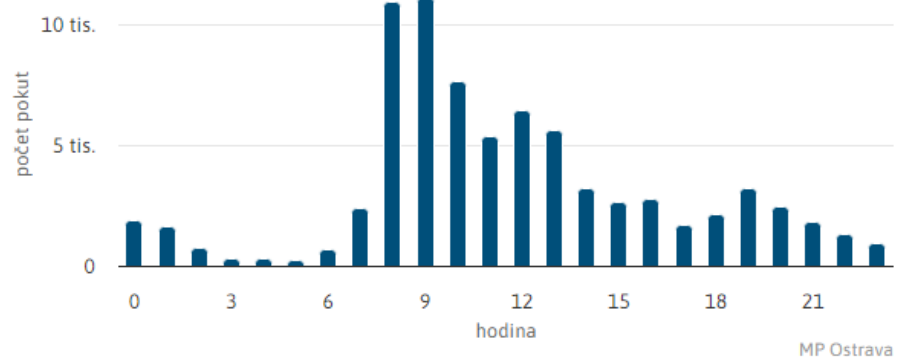


Srovnání volební účasti v Česku a zahraničí

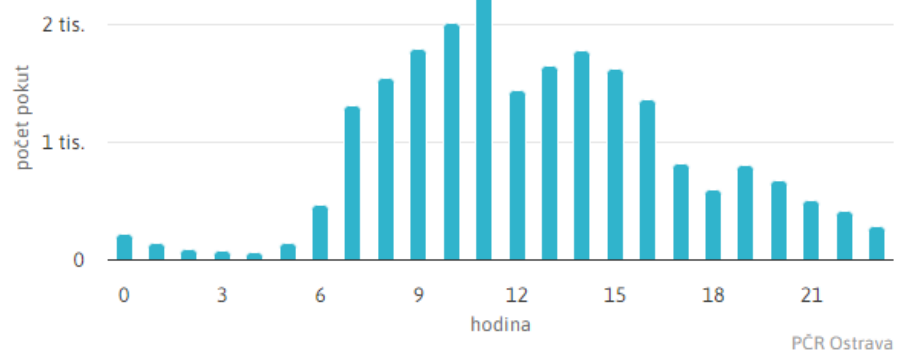
V roce 2002 byl rozdíl mezi volební účastí Čechů v zahraničí a doma 33 procentních bodů. Postupně se zmenšil až na 5 p.b. v posledních sněmovních volbách. Důvodem může být rostoucí počet Čechů v zahraničí - jejich volební chování tak má blíž k většinové populaci.



Pokuty během dne, MP

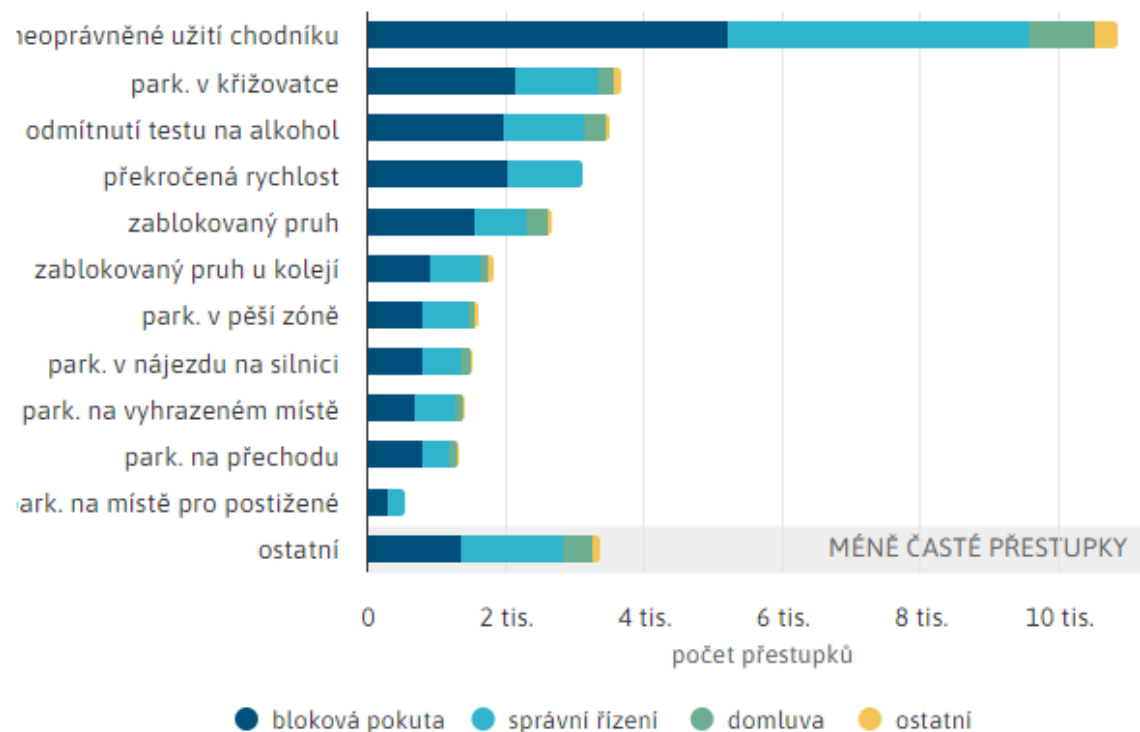


Pokuty během dne, PČR



Dopravní přestupky MP Brno 2022

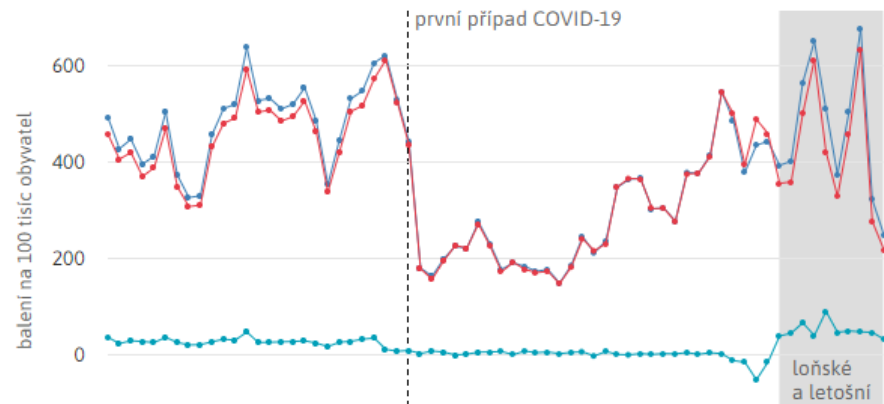
Nejčastějším konkrétním přestupkem je *neoprávněné užití chodníku* s téměř 11 tisíci záznamy. Z grafu ale vynecháváme 51 tisíc (59 % ze všech) přestupků pod obecným paragrafem, který indikuje porušení značky.



Zdroj: MP Brno

Měsíční předpisy a výdeje penicilinových antibiotik

Před pandemií zůstávalo nevyzvednutých kolem 5 % receptů. Loni na podzim číslo překonalo 10 %, letos v březnu propadly recepty na 17 % penicilinových antibiotik. Příčinou byl a je výpadek léků.



Změna životního stylu zachránila víc životů než léčba

Mezi lety 1985 a 2007 se počet úmrtí na ischemickou chorobu srdeční (neboli srdeční infarkt) snížil o 12 tisíc. 52 % životů zachránila lepší životospráva, 43 % léčba, 5 % zůstává nevysvětlených.

● 100 zachráněných životů ● 100 zmařených životů

Zlepšení rizikových faktorů zachránilo 6330 životů, má 52% podíl na celkovém snížení úmrtnosti

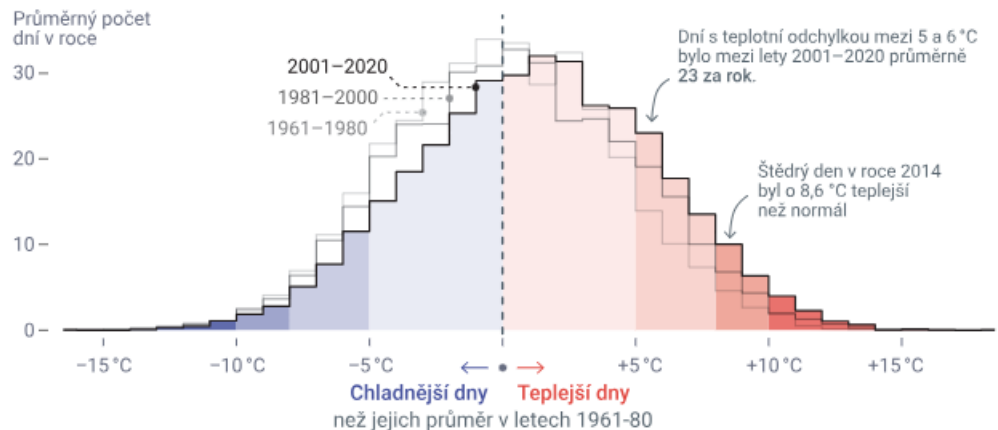


Léky a lékařské zákroky zachránily 5220 lidí, na nižší úmrtnosti měly 43% podíl



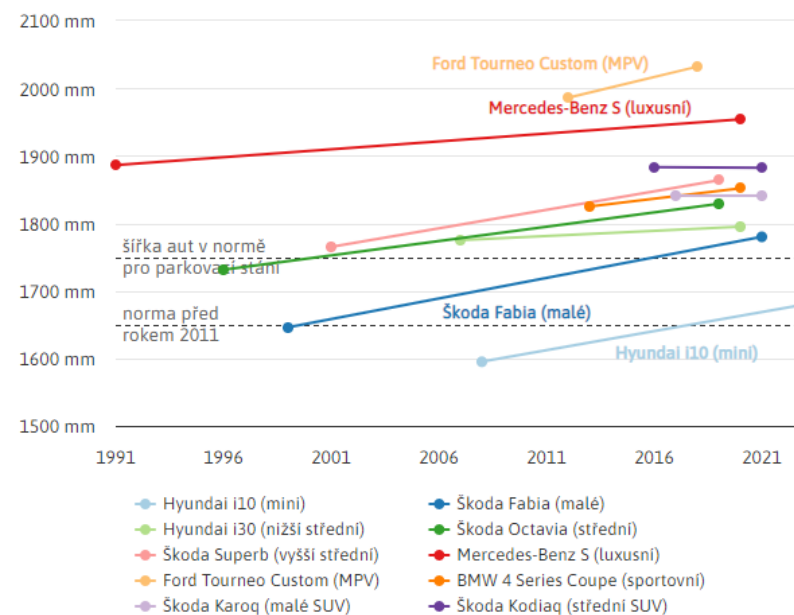
Zdroj dat: studie Explaining the decline in coronary heart disease | Vizualizace: iROZHLAS.cz, Jan Boček

VÝVOJ DENNÍCH TEPLOTNÍCH ODCHYLEK V ČR



🕒 Všechny dny v letech 1961–2020 📍 nejvyšší denní teploty 📅 referenční období 1961–1980

Šířka nejprodávanějších modelů



Zdroj: www.auto-data.net

Co je to statistika?

- Způsob **získání odpovědi** na (**výzkumnou**) otázku na základě **číselných dat**
 - Jak přítomnost památníků 2. sv. války ovlivňuje výsledky stran?
 - Jaké jsou vlastnosti voličů Andreje Babiše?
 - Jaké narativy obsahují o covidu obsahovaly vybrané dezinformační platformy?
 - Jak spolu souvisí nespokojenost se stranami a volební chování?
- „vědecká metoda“
 - Sama o sobě
 - Doplněk jiných metod
 - Experimenty
 - komparace

Statistika

- *„It is a key component in the scientific toolbox and one of the main ways we have of describing the natural world and of finding out how it works.“ (Hector 2021)*
- *„Statistics is a bit like sticking your finger into a revolving fan blade: sometimes it’s very painful, but it does give you the power to answer interesting questions.“ (Field 2009)*

Jak statistika funguje?

- Hledání vzorců ve velkém množství případů



“Without data you’re just another person with an opinion”

- W. Edwards Deming

- Souhrn technik, pomocí kterých se realizuje komplexní práce s daty:
 - Sběr
 - **Analýza**
 - Interpretace
 - *(a vizualizace)*
- Vysoká formalizace
 - Od dat k výsledku vede jasná cesta
 - Stejná pro jakéhokoli výzkumníka
 - U kvalitativních metod hraje velkou roli osobnost výzkumníka

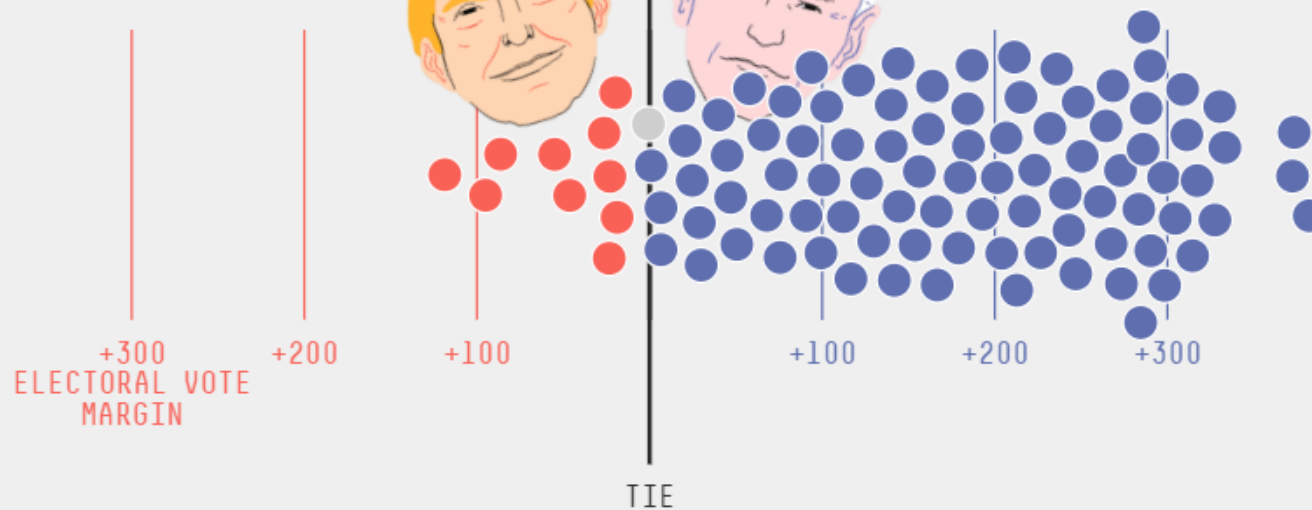
Není to pouze teorie

Biden is *favored* to win the election

We simulate the election 40,000 times to see who wins most often. The sample of 100 outcomes below gives you a good idea of the range of scenarios our model thinks is possible.

Trump wins
10 in 100

Biden wins
89 in 100



FiveThirtyEight

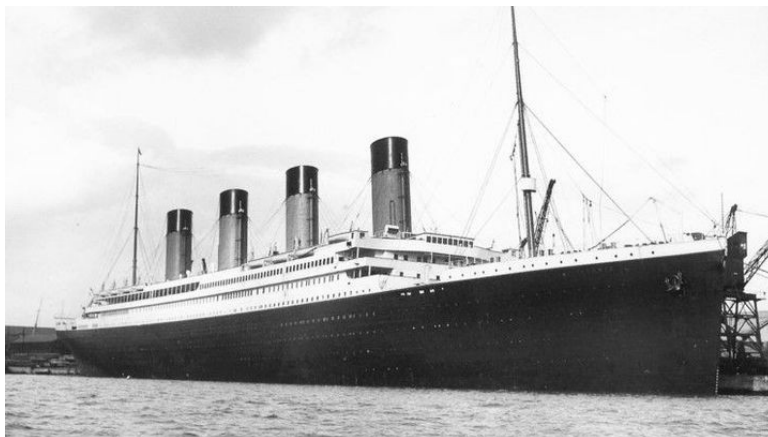


- Trump win
- Biden win
- No Electoral College majority, House decides election

Způsob vzniku dat

- Observační
 - Pozorujeme konání aktérů a zaznamenáváme do tabulky
 - Hlasování v parlamentu, prezidentské projevy, billboardy u silnice, pasažéři na Titanicu
- Dotazníková data
 - Zeptáme se sady respondentů na otázky
 - <https://www.surveymonkey.com/r/CJP8DC6>

Survived	Pclass	Name	Sex	Age	Siblings/Spouses Aboard	Parents/Children Aboard
0	3	Mr. Owen Harris Braund	male	22	1	0
1	1	Mrs. John Bradley (Florence Briggs Thayer) Cumings	female	38	1	0
1	3	Miss. Laina Heikkinen	female	26	0	0
1	1	Mrs. Jacques Heath (Lily May Peel) Futrelle	female	35	1	0
0	3	Mr. William Henry Allen	male	35	0	0
0	3	Mr. James Moran	male	27	0	0
0	1	Mr. Timothy J McCarthy	male	54	0	0
0	3	Master. Gosta Leonard Palsson	male	2	3	1
1	3	Mrs. Oscar W (Elisabeth Vilhelmina Berg) Johnson	female	27	0	2
1	2	Mrs. Nicholas (Adele Achem) Nasser	female	14	1	0
1	3	Miss. Marguerite Rut Sandstrom	female	4	1	1
1	1	Miss. Elizabeth Bonnell	female	58	0	0
0	3	Mr. William Henry Saundercock	male	20	0	0
0	3	Mr. Anders Johan Andersson	male	39	1	5
0	3	Miss. Hulda Amanda Adolfina Vestrom	female	14	0	0
1	2	Mrs. (Mary D Kingcome) Hewlett	female	55	0	0
0	3	Master. Eugene Rice	male	2	4	1
1	2	Mr. Charles Eugene Williams	male	23	0	0
0	3	Mrs. Julius (Emelia Maria Vandemoortele) Vander Planke	female	31	1	0
1	3	Mrs. Fatima Masselmani	female	22	0	0
0	2	Mr. Joseph J Fynney	male	35	0	0
1	2	Mr. Lawrence Beesley	male	34	0	0
1	3	Miss. Anna McGowan	female	15	0	0
1	1	Mr. William Thompson Sloper	male	28	0	0
0	3	Miss. Torborg Danira Palsson	female	8	3	1
1	3	Mrs. Carl Oscar (Selma Augusta Emilia Johansson) Asplund	female	38	1	5
0	3	Mr. Farred Chehab Emir	male	26	0	0
0	1	Mr. Charles Alexander Fortune	male	19	3	2
1	3	Miss. Ellen O'Dwyer	female	24	0	0
0	3	Mr. Lallo Todoroff	male	23	0	0
0	1	Don. Manuel E Uruchurtu	male	40	0	0
1	1	Mrs. William Augustus (Marie Eugenie) Spencer	female	48	1	0
1	3	Miss. Mary Agatha Glynn	female	18	0	0
0	2	Mr. Edward H Wheadon	male	66	0	0
0	1	Mr. Edgar Joseph Meyer	male	28	1	0
0	1	Mr. Alexander Oskar Holverson	male	42	1	0
1	3	Mr. Hanna Mamee	male	18	0	0
0	3	Mr. Ernest Charles Cann	male	21	0	0
0	3	Miss. Augusta Maria Vander Planke	female	18	2	0
1	3	Miss. Jamila Nicola-Yarred	female	14	1	0



No. 18,978 "WORLD'S LARGEST" TUESDAY, APRIL 16, 1912, 38 PAGES PRICE ONE CENT "WORLD'S LARGEST"

J. J. ASTOR LOST ON TITANIC

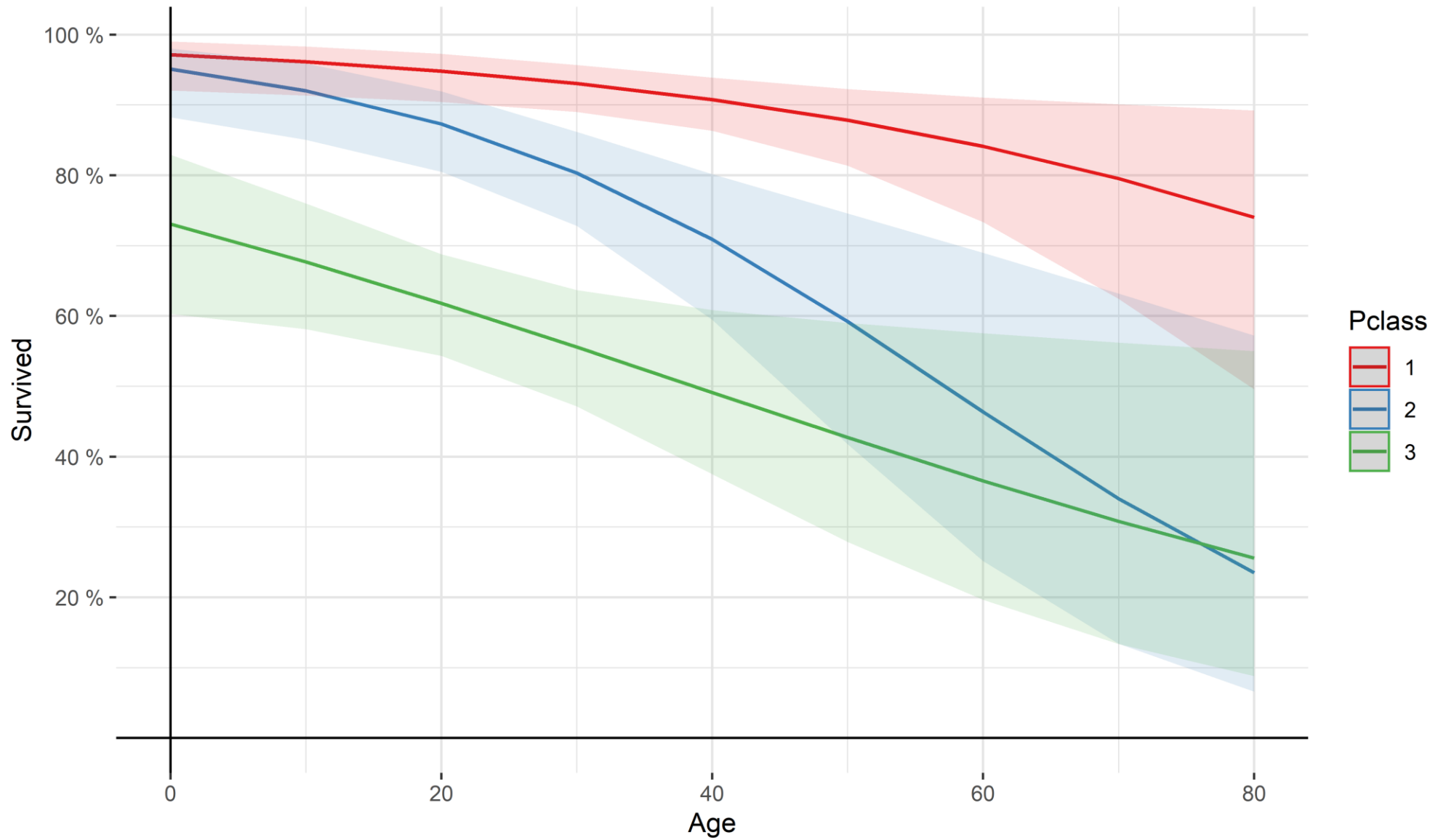
1,500 TO 1,800 DEAD

John Jacob Astor was among the passengers who went down with the ship, according to a wireless dispatch received by Bradstreets last night from the liner Olympic. Mrs. Astor was saved and is being brought to shore by the Carpathia.

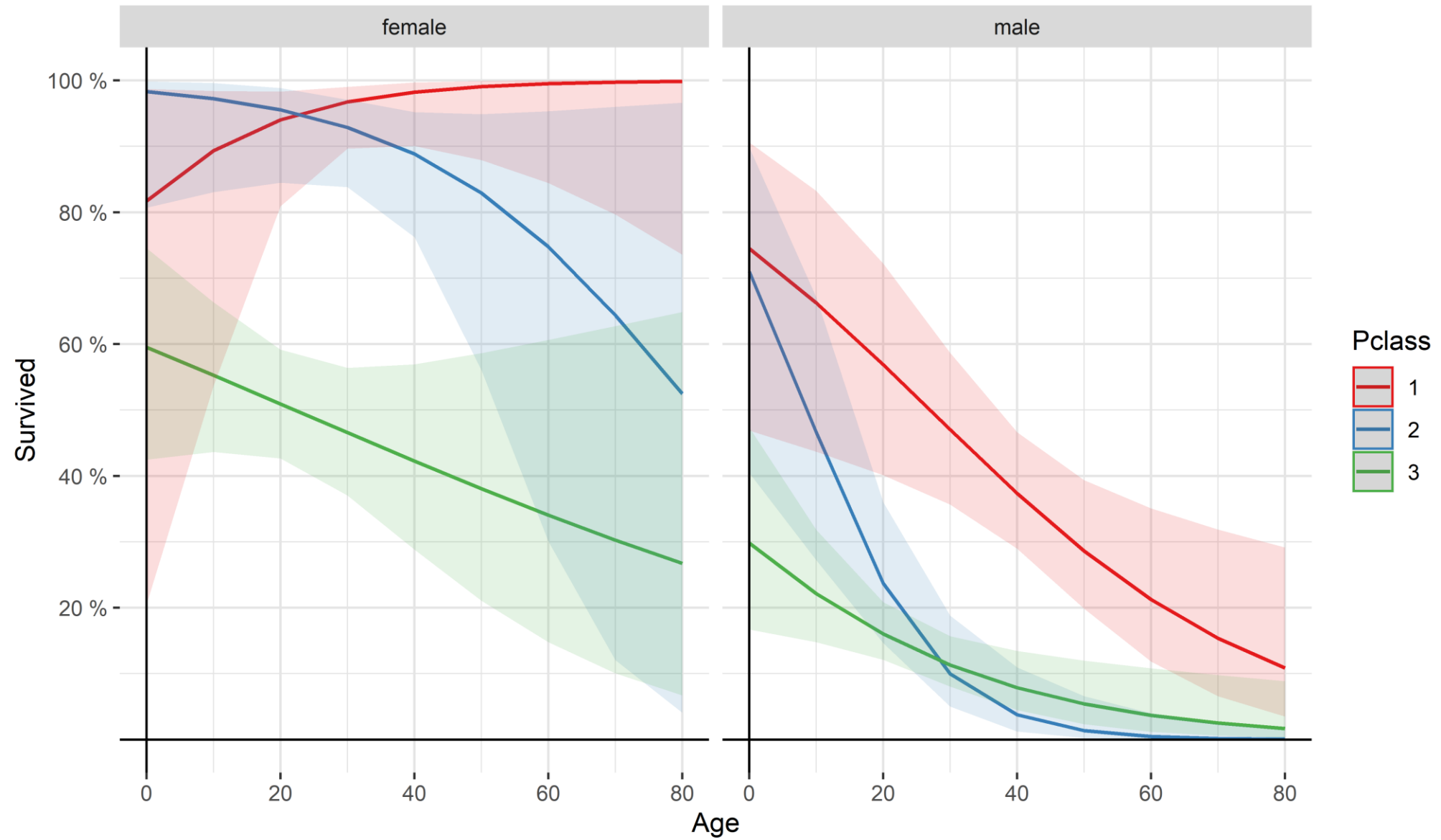
The Wireless Operator at Cape Race, Newfoundland, Flashes: "Eighteen Hundred Lives Have Been Lost in the Wreck of the Titanic."

- Data nám umožňují odpovídat různé otázky:
- Jaký měl vliv věk, pohlaví a cestovní třída vliv na šance na přežití?

Predicted probabilities of Survived



Predicted probabilities of Survived



Jak to funguje (Spiegelhalter 2019)

- Problem – Plan – Data – Analysis - Conclusion
1. Na základě pozorování si stanovíme problém
 2. Naplánujeme si postup
 3. (Stanovíme si teorii vysvětlující náš problém -> hypotézy)
 4. Získáme vhodná data a otestujeme platnost hypotézy
 5. Vytvoříme důvěryhodný závěr z našich analýz

Jak to funguje (Spiegelhalter 2019)

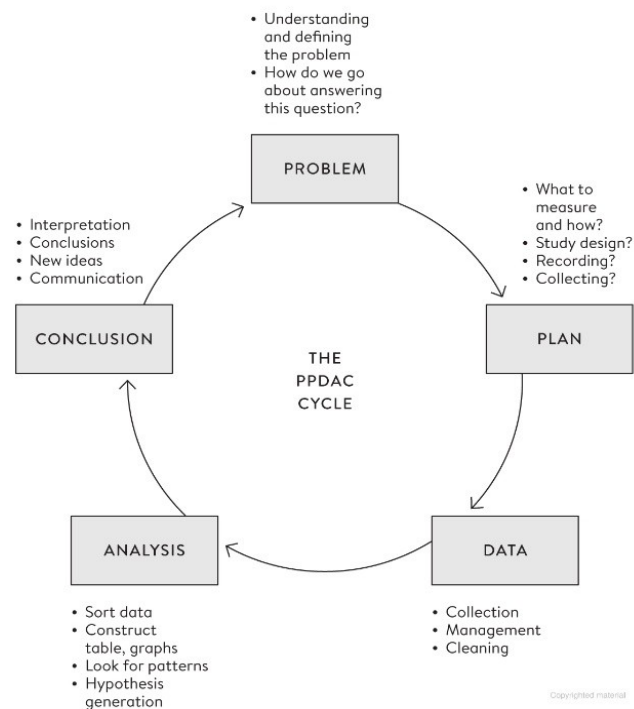


Figure 0.3

The PPDAC problem-solving cycle, going from Problem, Plan, Data, Analysis to Conclusion and communication, and starting again on another cycle.

Data: proměnné a případy

- Statistická analýza bez (velkého počtu) dat není možná
- Potřebujeme vědět:
 - Co měříme (koncept)
 - Jak to měříme (operacionalizace)
 - Na jaké úrovni měříme (jednotlivci, regiony, státy)
 - Co je jednotka analýzy (volič, politická strana, obec)

Rychlost auta

- Jak měřit: maximální rychlost auta v km/hod (operacionalizace)
 - Úroveň měření: individuální
 - Jednotka analýzy: auto
-
- Jednoduché a vcelku nesporné



Spokojenost s demokracií

- Jak měřit?
- Úroveň měření? (občané vs. státy)
- Občané:
 - Jak spokojeni jste se stavem demokracie ve vaší zemi? (škála 1-10)
 - Souhlasíte s následujícím tvrzením? Liberální demokracie s pravidelnými volbami je vhodnější než vláda silného lídra. (ano/ne)
- Odlišné způsoby měření → odlišné proměnné

Proč to nemíchat

- Úroveň měření vs. interpretace dosažených zjištění
- Ekologická chyba – aplikace výsledků na určitou úroveň bez toho, aby k tomu opravňovala podoba dat a analýzy
- Příklad:
 - Zjištění: konzervativní strany mají vyšší podporu v obcích s vyšší mírou nezaměstnanosti
 - Interpretace: nezaměstnaní lidé volí konzervativní strany
 - **Kde je chyba?**

Proměnná

- Způsob uchopení jevů, které zkoumáme
- „měření“ konceptů
- Její hodnota variuje (jinak jde o konstantu)

- Klíčový prvek nejen pro kvantitativní výzkum

- Denní teplota, počet hlasu strany ve volbách, cena ovoce, e-mailové adresy zaměstnanců univerzity

Dělení proměnných

- Dle postavení ve výzkumu:
 - Nezávislá (nezávisle) proměnná – předpokládaná příčina
 - Závislá (závisle) proměnná – předpokládaný následek
- Předpoklad kauzálního vztahu NP → ZP
- Základ pro formulaci hypotéz
- Přítomnost třetích proměnných a jejich kontrola

Dělení proměnných dle úrovně měření

		Možnost seřazení hodnot	Stejně rozdíly hodnot	Věcně významná nula
Kategorické	Nominální	-	-	-
	Ordinální	+	-	-
Kardinální	Intervalové	+	+	-
	Poměrové	+	+	+

- Kardinální:
 - Diskrétní – pouze celá čísla
 - Spojitá – teoreticky nekonečné množství hodnot

Příklady proměnných

- Nominální:
 - jména osob, názvy obcí
 - pouze dvě hodnoty – binární (dummy) – student (ano/ne), účast ve volbách (ano/ne)
- Ordinální – stupně vzdělání, vojenské hodnosti
- Intervalová – teplota v st. Celsia, poschodí v domě
- Poměrové – počet členů domácnosti, míra nezaměstnanosti
- **Důležitý je obsah, ne formální zápis** – jména osob se dají seřadit abecedně, ale nejde o věcně smysluplné pořadí (stejně tak telefonní čísla, adresy...)

Klíčové pojmy na závěr

- Reliabilita:

- Schopnost dosažení stejných výsledků při opakovaném měření
- Jak spolehlivá je naše operacionalizace a naše měření?
- Bude stejná hodnota naměřena zítra?
- Když kategorizací a měřením pověřím dva různé lidi, naměří stejnou hodnotu?

- Validita:

- Schopnost měření toho, co skutečně chceme měřit
- Schopnost generalizace našich zjištění (externí validita) – jiná přednáška

Validita a reliabilita (Lundberg 2006)

