

Začínáme s čištěním dat v programe Excel (Power Query)

Zlaté pravidlo: Nikdy nepřepisujte ani ručně neupravujte původní nezpracovaná data. Každé rozhodnutí týkající se čištění dat zaznamenejte.

Pro čištění a přípravu výzkumných dat existuje několik možností – nejčastěji se používají **Excel**, **R** nebo **Python**. Výběr závisí především na oborových zvyklostech a dostupných dovednostech týmu. Každý nástroj má své výhody a omezení, a proto je důležité vybrat ten, který nejlépe odpovídá potřebám vašeho projektu.

Excel je nejrozšířenější tabulkový procesor, který nabízí několik úrovní práce s daty. Na základní úrovni umožňuje manuální úpravy buněk, filtrování, řazení a použití jednoduchých vzorců přímo v tabulce. Pro pokročilejší uživatele nabízí rozšířené možnosti prostřednictvím **VBA (Visual Basic for Applications)** a **maker**, která umožňují automatizovat opakované úkony.

Nástroj Power Query je součástí Excelu a v novějších verzích integrován přímo v menu Data. Na rozdíl od manuálních úprav, Power Query automaticky zaznamenává každý transformační krok do přehledného panelu „Aplikované kroky“. To znamená, že celý pracovní postup je zdokumentován, transparentní a lze jej kdykoli znovu spustit z původních dat, což jsou klíčové principy reprodukovatelného výzkumu.

Proč je čištění dat důležité

Výzkumná data jsou jen málokdy připravena k analýze bez dalších úprav. Soubory mohou obsahovat nekonzistentní názvy proměnných, různé formáty data, chybějící hodnoty, duplicitní záznamy, neočekávané kategorie, skryté mezery, nesprávné datové typy nebo hodnoty, které nejsou fyzikálně možné. Tyto problémy mohou způsobovat chyby ve statistických modelech, dále komplikovat slučování datových souborů a ztěžovat ostatním výzkumníkům pochopení nebo opětovné využití výzkumných dat.

Čištění dat proto není pouze přípravným krokem. Je součástí zajištění kvality výzkumu. Dobře naplánovaný postup čištění dat zajišťuje, že transformace nezpracovaných dat na zpracovaná data probíhá transparentně, kontrolovatelně a reprodukovatelně. Zároveň podporuje principy FAIR tím, že zlepšuje dohledatelnost, dostupnost, interoperabilitu a opětovnou využitelnost výzkumných dat.

Před samotnou prací v programe Excel:

Přehledná struktura projektu pomáhá předcházet nechtěnému přepsání souborů a usnadňuje spolupracovníkům, recenzentům i vám samotným v budoucnu orientaci v analýze. Uchovávejte nezpracovaná data, zpracovaná data, skripty, dokumentaci a výstupy v samostatných složkách.

Pojmenovávání: Používejte popisné názvy souborů, vyhněte se mezerám a speciálním znakům a v případě, že je součástí názvu souboru datum, používejte formát YYYYMMDD. Například: floodplain_tree_growth_20260727_v01.csv.

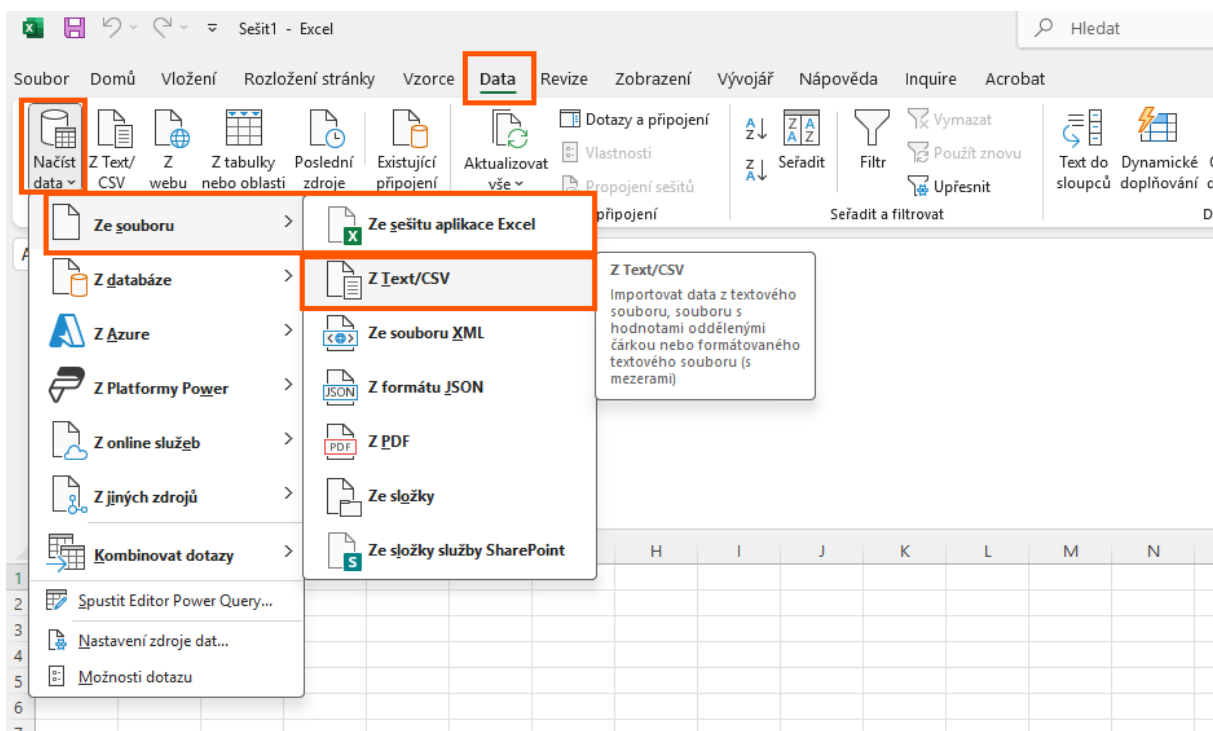
Zálohování: Dodržujte princip 3-2-1: uchovávejte tři kopie dat alespoň na dvou různých typech úložišť, přičemž jedna kopie by měla být uložena mimo pracoviště nebo ve schváleném institucionálním systému.

1. Co je PowerQuery a jak ho spustit

Power Query je nástroj na transformaci a čištění dat, který je součástí Excelu, takže není potřeba instalovat žádný další program. Jelikož je prostředí vizuálně podobné Excelu, je intuitivnější pro uživatele, kteří již pracují s tabulkovými procesory, ale na rozdíl od běžného Excelu automaticky zaznamenává všechny provedené úpravy.

2. Importujte data, aniž byste měnili zdrojový soubor

1. Otevřete Excel
2. Klikněte na **Data** → **Načíst data** → **Ze souboru**
3. Vyberte typ souboru:
 - **Ze sešitu Excelu** pro .xlsx
 - **Z textu/CSV** pro .csv
4. Najděte a vyberte soubor ve složce data_raw



5. Klikněte na **Importovat**
6. V okně **Navigátor** vyberte list nebo tabulku
7. **Neklikněte na Načíst** – klikněte na **Transformovat data**

Navigátor

☐ Vybrat více položek

Možnosti zobrazení ▾

▲ NATURE_messy (1).xlsx [1]

List1

List1

Column1	Nature project, photos are in the folder on Google Drive. Disturbance ..
updated: 01/05/2026	
Record ID	Plot ID
REC-0001	NO-11
REC-0002	RI-10
REC-0003	OL-08
REC-0004	OL-12
REC-0005	OA-09
REC-0006	MO-12
REC-0007	OL-08
REC-0008	MO-06
REC-0009	RI-03
REC-0010	OA-05
REC-0011	NO-07
REC-0012	OL-03
REC-0013	OL-09
REC-0014	RI-09
REC-0015	NO-11
REC-0016	NO-03
REC-0017	MO-08
REC-0018	OL-04
REC-0019	RI-12
REC-0020	PI-01
REC-0021	MO-05

Načíst ▾ Transformovat data Zrušit

Důležité: Pokud se formáty liší (oddělovače, jazyk a speciální znaky), nastavte je při importu CSV:

- Nastavte **Oddělovač** (středník, tečka, čárka...)
- Nastavte **Původ souboru** (pomůže s lokalizací)

□ ×

NATURE_messy (1).csv

Původ souboru
1250: Střední Evropa (Windows)

Oddělovač
Středník

Zjišťování datového typu
Založené na prvních 200 řádcích

Column1	Column2	Column3	Column4	Column5	Column6	Column7	
	Nature project, photos are in the folder on Google Driv...						
updated: 01/05/2026							
Record ID	Plot ID	Site Name	Researcher	Observation Date	Visit time	Taxon group	Sp
REC-0001	NO-11	north fen	Tomáš Blaha?	09/19/2025	morning	BIRD	Pe
REC-0002	RI-10	Riverbend	Martin Šimek	2025/04/03	07:00	Avian	Fr
REC-0003	OL-08	Old Quarry	L. Horák	2025/05/07	19:15	Bugs	Bo
REC-0004	OL-12	old quarry	A. Nováková	2025-28-05	10:00	insect	Bo
REC-0005	OA-09	Oak Ridge	A. Nováková	04/25/2025	11:30	Amphibia	Li
REC-0006	MO-12	Moss Hollow	L. Horak	Apr 07, 2025	08:00	plants	Qu
REC-0007	OL-08	Old Quarry	A. Novakova	Sep 08, 2025	14:15	insect	Ca
REC-0008	MO-06	Moss Hollow	A. Novakova	20.05.2025	07:00	insect	Aq
REC-0009	RI-03	River Bend	L. Horák	2025-23-08	13:15	birds	Pe
REC-0010	OA-05	Oak Ridge	Tomáš Blaha	30.09.2025	08:45	Flora	Ar
REC-0011	NO-07	north fen	Hana P.?	09/16/2025	7:5	Amphibia	Bu
REC-0012	OL-03	Old Quarry	L. Horák	2025/05/07	14:30	Amphibian	Li
REC-0013	OL-09	Old Quarry	A. Nováková	2025-10-17	15:45	Fungi	Ar
REC-0014	RI-09	Riverbend	Hana P.	Apr 22, 2025	06:15	Amphibian	Li
REC-0015	NO-11	North Fen	Martin Šimek	20.03.2025	7:5	Fungus	Fo
REC-0016	NO-03	North-Fen	A. Nováková	2025/04/25	08:00	Arthropod	Pl
REC-0017	MO-08	moss hollow	Hana P.	08/27/2025	13:30	Plant	Ca

Načíst Transformovat data Zrušit

3. Zkontrolujte data před čištěním

Při kontrole dat je dobré všimnout si, co PowerQuery skutečně nainportoval. Nikdy automaticky nepředpokládejte, že sloupec obsahující čísla byl načten jako číselný datový typ nebo že data byla správně interpretována jako datum. Zkontrolujte jak strukturu dat, tak samotné hodnoty.

1. Otevře se editor Power Query
2. Prohlédněte si náhled dat v hlavní části okna
3. Zkontrolujte ikony vedle názvů sloupců (označují datový typ):
 - **123** = číslo
 - **ABC** = text
 - **Kalendář** = datum

Klikněte na záložku „Zobrazení“ a zaškrtněte možnosti „Zobrazit prázdné znaky“, „Kvalita sloupce“, „Distribuce sloupce“ a „Profil sloupce“:

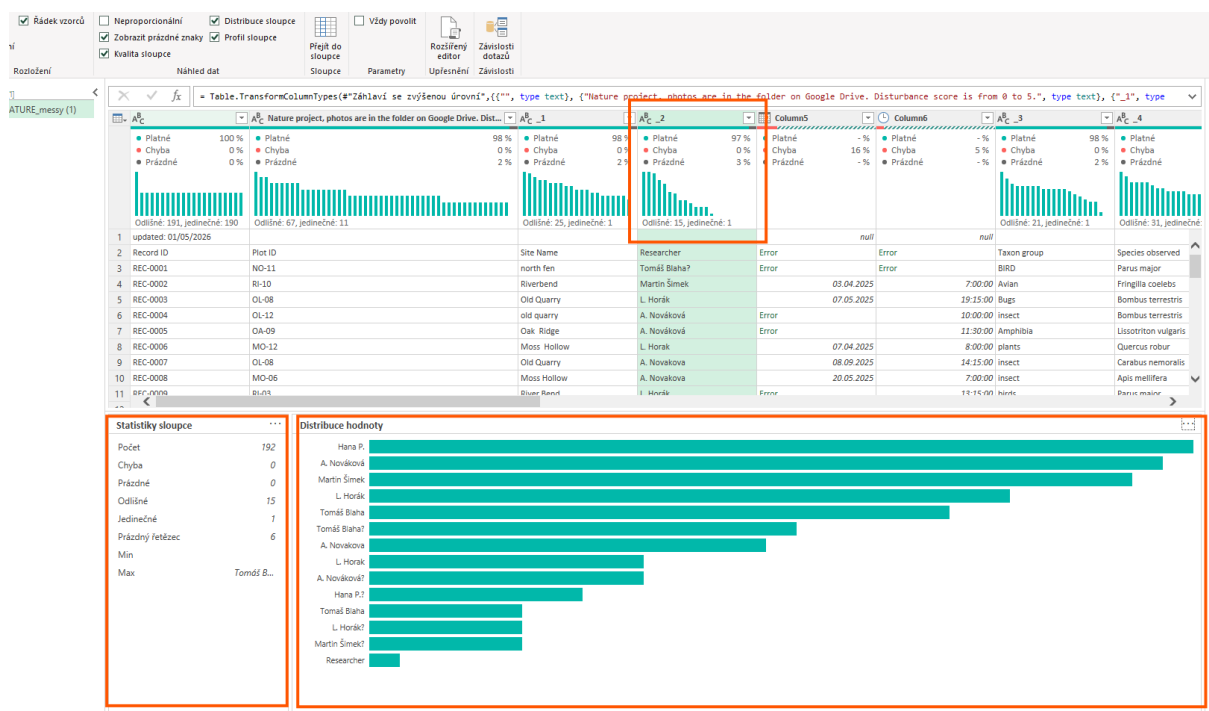
Dotazy [1]

NATURE_messy (1)

Table.TransformColumnTypes(#"Záhlaví se zvýšenou úrovní",{{"", type text}, {"Nature project, photos are in the folder on Google Drive. Dist...", type text}, {"A^B_C _1", type text}, {"A^B_C _2", type text})

	A^B_C	A^B_C Nature project, photos are in the folder on Google Drive. Dist...	A^B_C _1	A^B_C _2
1	updated: 01/05/2026			
2	Record ID	Plot ID	Site Name	Researcher
3	REC-0001	NO-11	north fen	Tomáš B
4	REC-0002	RI-10	Riverbend	Martin Š
5	REC-0003	OL-08	Old Quarry	L. Horák
6	REC-0004	OL-12	old quarry	A. Novál
7	REC-0005	OA-09	Oak Ridge	A. Novál
8	REC-0006	MO-12	Moss Hollow	L. Horák
9	REC-0007	OL-08	Old Quarry	A. Novál

Pak je možné lépe posoudit množství chybějících hodnot nebo statistiky sloupců:



Při kontrole dat se zaměřte na:

- chybějící hodnoty zaznamenané pomocí několika různých kódů;
- duplicitní řádky nebo opakující se identifikátory;
- neočekávané rozdíly v pravopisu nebo používání velkých a malých písmen v kategoriích;
- nemožné hodnoty, například záporný průměr stromu nebo datum mimo období výzkumu;
- chybějící, smíšené nebo nezdokumentované jednotky;
- řádky obsahující poznámky, mezisoučty nebo metadata namísto jednotlivých pozorování.

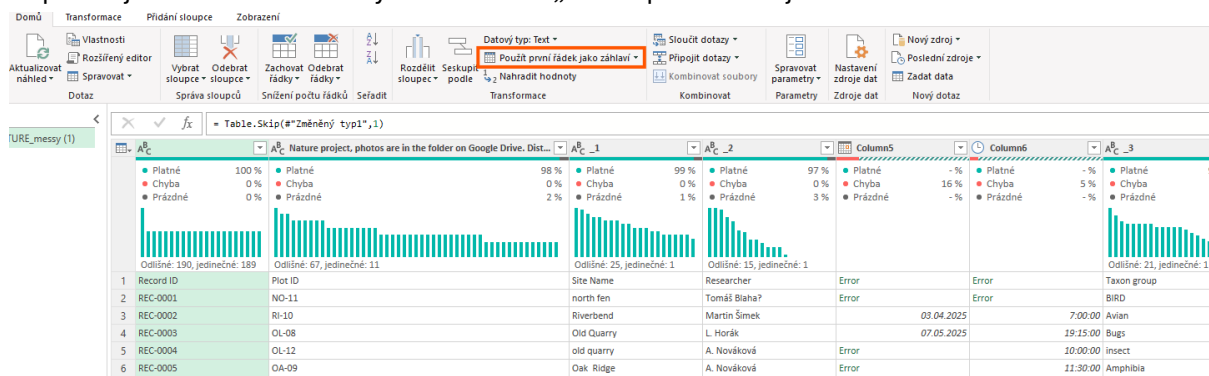
4. Standardizujte názvy a datové typy

Konzistentní názvy a datové typy jsou nezbytné pro spolehlivou analýzu. Názvy proměnných by měly být stručné, ale výstižné, měly by používat jednotnou konvenci pojmenování a neměly by obsahovat mezery

Názvy sloupců:

1. Dvakrát klikněte na název sloupce
2. Napište nový název (bez mezer, malá písmena, podtržítka místo mezer)
3. Stiskněte **Enter**
4. Nebo: Klikněte pravým tlačítkem na sloupec → **Přejmenovat**

Případně je možné kliknout a vybrat možnost „Použít první řádek jako záhlaví“:



Při kliknutí na ikonu vedle sloupce je možné změnit datový typ. Zkontrolujte a případně změňte na správný datový typ všechny sloupce.

Vyberte správný typ:

- **Celé číslo** pro čísla bez desetinných míst
- **Desetinné číslo** pro čísla s desetinnými místy
- **Text** pro textové hodnoty
- **Datum** pro data
- **Logická hodnota** pro TRUE/FALSE

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8
1	Nature project, photos are in the folder on Google Drive. Disturbance...						
2	updated: 01/05/2026						
3	Record ID	Plot ID	Site Name	Researcher			
4	REC-0001	NO-11	north fen	Tomáš Blaha?			
5	REC-0002	RI-10	Riverbend	Martin Šimek			
6	REC-0003	OL-08	Old Quarry	L. Horák			
7	REC-0004	OL-12	old quarry	A. Nováková			
8	REC-0005	OA-09	Oak Ridge	A. Nováková			
9	REC-0006	MO-12	Moss Hollow	L. Horák			
10	REC-0007	OL-08	Old Quarry	A. Nováková			
11	REC-0008	MO-06	Moss Hollow	A. Nováková			
12	REC-0009	RI-03	River Bend	L. Horák			
13	REC-0010	OA-05	Oak Ridge	Tomáš Blaha			
14	REC-0011	NO-07	north fen	Hana P.?			
15	REC-0012	OL-03	Old Quarry	L. Horák			
16	REC-0013	OL-09	Old Quarry	A. Nováková			
17	REC-0014	RI-09	Riverbend	Hana P.			
18	REC-0015	NO-11	North Fen	Martin Šimek			
19	REC-0016	NO-03	North-Fen	A. Nováková			
20	REC-0017	MO-08	moss hollow	Hana P.			
21	REC-0018	OL-04	Old quarry	Tomáš Blaha			
22	REC-0019	RI-12	river bend	A. Nováková			
23	REC-0020	PI-01	Pine transect	L. Horák			
24	REC-0021	MO-05	moss hollow	Martin Šimek			
25	REC-0022	MO-09	Moss Hollow	A. Nováková			
26	REC-0023	OL-05	Old quarry	Hana P.			
27	REC-0024	OA-10	Oak Ridge	Hana P.			
28	REC-0025	RI-04	R. Bend	Hana P.?			

Výsledek vždy zkontrolujte, protože při převodu mohou vzniknout chybějící hodnoty, pokud některou hodnotu nelze správně převést do požadovaného formátu.

1. Pokud se objeví sloupec **Chyby**, klikněte na jeho hlavičku
2. Podívejte se, které řádky se nepodařilo převést
3. Zaznamenejte počet chyb do dokumentace
4. Rozhodněte se, zda tyto řádky opravit nebo odstranit

Chyba u číselného řádku:

1.2 Column13	
Platné	- %
Chyba	3 %
Prázdné	- %
Error	59,2
	82
	53,5
	50,7
	44,4
	27,4
	10,5
	11,6
	0,72
23	
Error	65,6
	43,4
	57,6

Příklad: Po kliknutí na chybu se zobrazí informace, že hodnotu „Approx. 40“ se nepodařilo převést na číslo.



= #"Odebrané horní řádky"{11}[Column13]

 DataFormat.Error: Hodnotu se nám nepovedlo převést na typ Number.
Podrobnosti:
approx. 40

5. Vybírejte, filtrujte, řadte a měňte strukturu dat

Výběrem podmnožiny dat vytvoříte přehlednější pracovní tabulku a zároveň zachováte kompletní importovaný datový soubor. Místo ručního mazání řádků nebo sloupců používejte explicitně zapsaný kód.

Výběr sloupců:

1. Podržte Ctrl a klikněte na sloupce, které chcete ponechat
2. Klikněte pravým tlačítkem na jeden z vybraných sloupců
3. Vyberte Odstranit další sloupce

The screenshot displays the Open Science Mendel software interface. At the top, there is a menu bar with options like 'Soubor', 'Domů', 'Transformace', 'Přidání sloupce', and 'Zobrazení'. Below the menu is a toolbar with various icons for data manipulation. The main area shows a data table with columns labeled 'A1', 'A2', 'A3', and 'A4'. A context menu is open over the 'A2' column, showing options such as 'Kopírovat', 'Odebrat sloupce', 'Přidat sloupce z příkladu...', 'Odebrat duplikáty', 'Odebrat chyby', 'Nahradit hodnoty...', 'Výplnit', 'Změnit typ', 'Transformovat', 'Sloučit sloupce', 'Vytvořit datový typ', 'Seskupit podle...', 'Převést sloupce na řádky', 'Převést ostatní sloupce na řádky', 'Převést jenom vybrané sloupce na řádky', and 'Přesunout'. The data table contains rows with various identifiers and values, including 'Record ID', 'Plot ID', 'Site Name', 'Researcher', 'Date', 'Time', 'Taxon group', and 'Species observed'.

Vytvoření nového sloupce:

1. Klikněte na záložku **Přidat sloupec** → **Vlastní sloupec**
2. Zadejte název nového sloupce
3. Do vzorce napište výraz (např. [diameter_cm] / 2)
4. Klikněte na **OK**

Přesunutí sloupců:

1. Klikněte na sloupec
2. Klikněte na **Transformovat** → vyberte **Přesunout** → Doleva / Doprava / Na začátek / Na konec
3. Případně je možné přetáhnout záhlaví.

Filtrování řádků:

1. Klikněte na šipku v hlavičce sloupce
2. Zrušte zaškrtnutí hodnot, které chcete vyloučit
3. Nebo použijte Číselné filtry / Textové filtry:
 - a. Větší než, Menší než, Rovno
 - b. Obsahuje, Začíná na, Končí na
4. Klikněte na OK

The screenshot shows the Open Science software interface. The top menu bar includes 'Soubor', 'Domu', 'Transformace', 'Přidání sloupce', and 'Zobrazení'. The 'Transformace' menu is open, showing options like 'Vlastnosti', 'Rozšířený editor', 'Spravovat', 'Výbrat sloupce', 'Odebrat sloupce', 'Zachovat řádky', 'Odebrat řádky', 'Seřadit', 'Rozdělit sloupec', 'Seskupit podle', 'Použít první řádek jako záhlaví', 'Nahradit hodnoty', 'Sloučit dotazy', 'Připojit dotazy', 'Kombinovat soubory', 'Spravovat parametry', 'Nastavení zdroje dat', 'Nový zdroj', and 'Poslední zdroj'. The main window displays a data table with columns 'Column13', 'Column14', 'Column16', and 'Column17'. A filter menu is open for 'Column14', showing a list of values to be excluded, with 'Seřadit sestupně' and 'Zrušit filtrování' options. The data table shows rows of data, including 'Water Temp C', 'Error', and 'Column16'.

Řazení v rámci sloupce:

1. Klikněte na šipku v hlavičce sloupce
2. Vyberte **Seřadit vzestupně** nebo **Seřadit sestupně**
3. Pro více úrovní: Podržte **Shift** a klikněte na další sloupce

The screenshot shows a data table with multiple columns. A dropdown menu is open for 'Column8', displaying sorting options: 'Seřadit vzestupně' (Sort ascending) and 'Seřadit sestupně' (Sort descending). The table includes data such as weather conditions, error rates, and habitat types across 31 rows.

Odstranění duplicit:

1. Vyberte sloupce, podle kterých se mají hledat duplicity (např. unikátní ID)
2. Klikněte na **Domů** → **Odebrat řádky** → **Odebrat duplicitu**
3. Potvrďte v dialogovém okně

The screenshot shows the 'Odebrat duplicitu' (Remove duplicates) dialog box in the Mendelion software. The dialog is open, and the user is selecting columns for duplicate removal. The background shows a data table with columns like Record ID, Plot ID, Researcher, and Site Name.

6. Identifikujte a řešte chybějící hodnoty

S chybějícími hodnotami je třeba zacházet podle toho, co znamenají a jaká analýza má být provedena. Odstranění všech řádků obsahujících chybějící hodnotu je ve většině případů nevhodné. Nejprve zjistěte, kde se chybějící hodnoty vyskytují, a poté rozhodněte, zda je ponechat, vyloučit, překódovat nebo nahradit pomocí imputace. Důvod každého takového rozhodnutí zaznamenejte.

Nalezení chybějících hodnot:

1. Klikněte na **Zobrazení** → **Kvalita sloupce**
2. Podívejte se na procento prázdných hodnot u každého sloupce
3. Klikněte na číslo prázdných hodnot pro filtrování

Možnosti řešení:

A) Ponechat chybějící hodnoty:

- Pokud je chybějící hodnota informativní
- Pokud by odstranění zkreslovalo výsledky

B) Odstranit řádky:

1. Klikněte na počet prázdných hodnot → „Odebrat prázdné hodnoty“.

C) Nahradit hodnotou (pouze pokud má nahrazení jednoznačný význam):

1. Vyberte sloupec
2. Klikněte na **Transformovat** → **Nahradit hodnoty**
3. **Hledaná hodnota:** null nebo nechte prázdné
4. **Nahradit hodnotou:** napište hodnotu (např. 0 nebo "Unknown")
5. Klikněte na **OK**

Příklad: Chybějící počet sadeb nahradte nulou pouze pokud chybějící hodnota znamená "nebylo nic pozorováno", nikoli "pozorování neproběhlo".

7. Standardizujte textové a kategoriální hodnoty

Drobné rozdíly v mezerách, pravopisu nebo používání velkých a malých písmen vedou ke vzniku samostatných kategorií. Textové řetězce proto vyčistěte ještě před jejich převodem na faktory nebo před výpočtem souhrnných statistik podle skupin.

Pokud je třeba sjednotit větší množství kategoriálních označení, používejte zdokumentovanou převodní tabulku. Tento postup je přehlednější a lépe kontrolovatelný.

Odstranění mezer:

1. Vyberte textový sloupec
2. Klikněte na **Transformovat** → **Formát** → **Oříznout**
3. Pro odstranění všech mezer: **Formát** → **Vyčistit**

Sjednocení velkých a malých písmen:

1. Vyberte textový sloupec
2. Klikněte na **Transformovat** → **Formát**
3. Vyberte:
 - **Malá písmena** pro všechna malá
 - **Velká písmena** pro všechna velká
 - **Velké první písmeno** pro první písmeno velké

The screenshot shows the Open Science software interface. The 'Transform' menu is open, and the 'Format' submenu is visible. The 'Format' submenu includes options for text formatting: 'malá písmena' (lowercase), 'VELKÁ PÍSMENA' (uppercase), 'Velké první písmeno každého slova' (first letter of each word uppercase), 'Oříznout' (trim), 'Vyčistit' (clean), 'Přidat předponu' (add prefix), and 'Přidat příponu' (add suffix). The main window displays a data table with columns for Plot ID, Site Name, Researcher, Observation Date, and Visit. The 'Nastavení dotazů' (Query Settings) panel on the right shows the 'Zdroj' (Source) as 'NATURE_messy (1)' and the 'Výběr' (Selection) as 'Změněný typ' (Changed type).

Nahrazení kategorií:

1. Podívejte se v části “Distribuce hodnoty” na hodnoty, které jsou stejné, ale jinak zapsané (u příkladu např. “river bend” a “r. bend”)
2. Vyberte sloupec
3. Klikněte na **Transformace** → **Nahradit hodnoty**
4. **Hledaná hodnota:** r. bend
5. **Nahradit hodnotou:** river bend
6. Klikněte na **OK**
7. Opakujte pro další kategorie

8. Převeďte a ověřte datумы

Převod textu na datum:

1. Vyberte sloupec s daty
2. Klikněte na ikonu datového typu vedle názvu sloupce
3. Vyberte **Datum**
4. Pokud formát není rozpoznán:
 - Klikněte na **Transformace** → **Datum a čas** → **Použít místní nastavení**
 - Klikněte na **OK**

Kontrola rozsahu datumů:

1. Klikněte na šipku v hlavičce sloupce
2. Podívejte se na nejmenší a největší datum ve filtru
3. Zkontrolujte, zda data spadají do věrohodného rozsahu.

Extrahování částí data:

1. Vyberte sloupec s datem
2. Klikněte na **Přidat sloupec** → **Datum**
3. Vyberte:
 - **Rok** pro extrakci roku
 - **Měsíc** pro extrakci měsíce
 - **Den** pro extrakci dne

The screenshot displays the Open Science software interface. The top menu bar includes 'Soubor', 'Domů', 'Transformace', 'Přidání sloupce', and 'Zobrazení'. The 'Přidání sloupce' menu is open, showing options like 'Rok', 'Měsíc', 'Čtvrtletí', 'Týden', and 'Den'. The main data table is visible with columns for 'Rok', 'Měsíc', 'Čtvrtletí', 'Týden', and 'Den'. The 'Rok' column is highlighted, and the 'Měsíc' column is also visible. The data table shows various records with dates and times, such as '03.04.2025 07:00' and '07.05.2025 19:15'. The interface also includes a sidebar with filters and a bottom status bar.

9. Identifikujte duplicity a ověřte záznamy

Přesně duplicitní záznam nemusí být vždy chybou a opakující se identifikátor může být v longitudinálních datech nebo při opakovaných měřeních zcela oprávněný. Nejprve proto určete, která kombinace proměnných má být v rámci dané studie jedinečná.

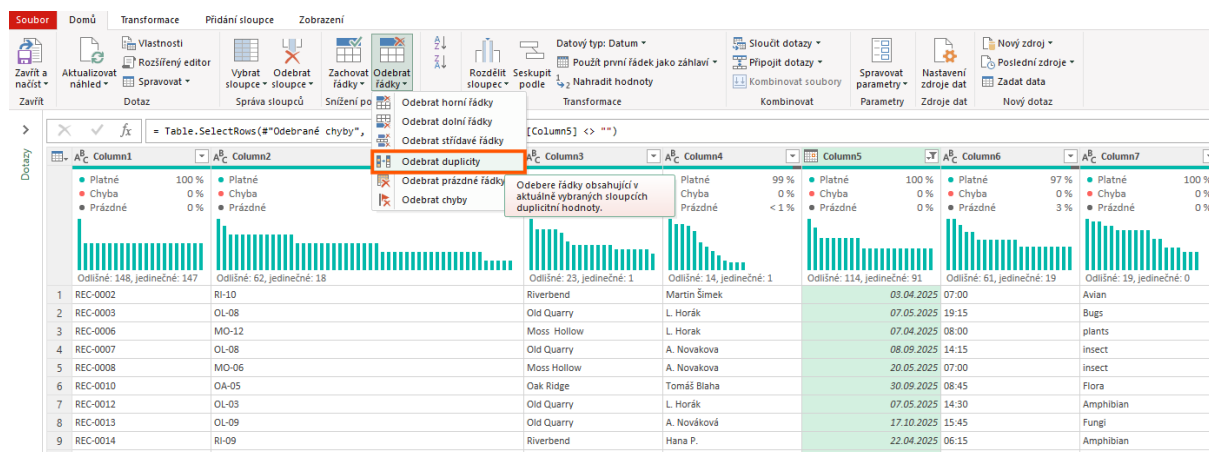
Ověřte také, zda hodnoty spadají do věrohodných rozsahů a zda kategoriální proměnné obsahují pouze povolené hodnoty. Podezřelé hodnoty raději označte ke kontrole, než abyste je bez zdokumentování odstranili.

Nalezení potenciálně duplicitních pozorování podle klíče:

1. Vytvořte skupinový dotaz: **Domů** → **Skupinové operace**
2. Seskupte podle: tree_id a sampling_date
3. Operace: **Počet řádků**
4. Filtrujte výsledky, kde počet > 1

Nalezení duplicit:

1. Vyberte sloupce, které by měly být jedinečné
2. Klikněte na **Domů** → **Odebrat řádky** → **Odebrat duplicity**
3. Power Query zobrazí počet odebraných duplicit



Ověření platných rozsahů:

1. Klikněte na **Přidat sloupec** → **Vlastní sloupec**
2. Název: kontrola_rozsah
3. Vzorec: if [diameter_cm] > 0 and [diameter_cm] < 300 then "OK" else "Zkontrolovat"
4. Klikněte na **OK**
5. Filtrujte na "Zkontrolovat" a ověřte hodnoty

Vytvoření příznaků kvality:

1. Vytvořte více vlastních sloupců pro různé kontroly
2. Příklad pro datum: `if [datum] >= #date(2000,1,1) and [datum] <= Date.From(DateTime.LocalNow()) then "OK" else "Zkontrolovat"`
3. Příklad pro status: `if [status] = "living" or [status] = "dead" or [status] = "unknown" then "OK" else "Zkontrolovat"`

10. Zkontrolujte vyčištěný datový soubor

Čištění dat je dokončeno až po důkladné kontrole výsledného datového souboru. Porovnejte počty řádků před a po čištění, zkontrolujte rozdělení hodnot a jednotlivé kategorie a ověřte, že všechny provedené transformace vedly k zamýšlenému výsledku.

Porovnání počtů řádků:

1. Podívejte se na spodní část editoru Power Query
2. Zobrazuje se počet řádků (např. "1000 řádků").
3. Porovnejte s původním počtem před čištěním

Kontrola rozdělení hodnot:

1. Klikněte na **Domů** → **Profil sloupce**
2. Prohlédněte si rozdělení hodnot u kategoriálních sloupců
3. Hledejte neočekávané kategorie

Souhrnné statistiky:

1. Klikněte pravým tlačítkem na číselný sloupec
2. Vyberte **Zobrazit souhrn**
3. Zobrazí se: Minimum, Maximum, Průměr, Počet

Vizuální kontrola:

1. Načtěte data do Excelu (**Domů** → **Zavřít a načíst**)
2. Vytvořte graf: **Vložení** → **Graf** → **Bodový**
3. Hledejte odlehlé hodnoty, asymetrická rozdělení, mezery v datech

11. Vytvořte přehledný pracovní postup

Panel Aplikované kroky:

1. V pravé části editoru je panel **Aplikované kroky**
2. Každá změna se automaticky zaznamená jako nový krok
3. Kliknutím na krok se vrátíte k dané fázi čištění
4. Kliknutím na **X** vedle kroku jej odstraníte

Přejmenování kroků pro dokumentaci:

1. Klikněte pravým tlačítkem na krok v panelu **Aplikované kroky**
2. Vyberte **Přejmenovat**
3. Napište popisný název (např. "Odebrat prázdné řádky", "Standardizovat názvy druhů")
4. Tím zdokumentujete, co každý krok dělá

12. Uložte výstupy a zdokumentujte pracovní postup

Export vyčištěných dat:

1. V editoru Power Query klikněte na **Domů** → **Zavřít a načíst**
2. Data se načtou do Excelu jako tabulka
3. Pro export do CSV:
 - Klikněte na **Soubor** → **Uložit jako**
 - Vyberte formát **CSV** (odděleno čárkou)
 - Uložte do složky data_processed

Obnovení dat:

1. Při aktualizaci zdrojových dat klikněte na **Data** → **Obnovit vše**
2. Power Query automaticky provede všechny uložené kroky čištění
3. Nová data se objeví ve stejném formátu

Vytvoření dokumentace:

1. V Excelu vytvořte list "Dokumentace"
2. Zaznamenejte:
 - Účel datového souboru
 - Zdroj dat a datum získání
 - Seznam všech transformačních kroků (z panelu Aplikované kroky)
 - Pravidla pro čištění a filtry
 - Verze Excelu a data poslední aktualizace
3. Uložte jako součást souboru