Visualization + Spatial + Base R : Side 1 of the Sheet

I. ggplot2 Basics

Chart Type	Geom	Notes
Barplot (raw)	geom_bar()	counts automatically
Barplot (summarized)	geom_bar()	after count()
Histogram	geom_histogram()	
Density	geom_density()	
Scatter	geom_point()	+ geom_smooth(method='lm') for regression

II. Bar charts positions:

- Stacked → default
- Side-by-side → position='dodge'
- Proportional → position='fill', color = outline, fill = inside

III. Spatial Data

- Earth ≈ ellipsoid
- CRS(Coordinate Reference System)/GCS(Geographic Coordinate System) = Datum + Projection
- Local CRSs → higher accuracy per region
- Always check CRS before combining datasets

IV. Packages

sf, terra, stars, elevatr, tidycensus, devtools::install_github("ropensci/USAboundaries"), tidyverse, dplyr

V. Base R / Lists / Data Frames

- Subset vector: x[1:3]
- First element of list: x[[1]]
- First 3 rows of df: x[1:3,]
- Pull column: tb$x, tb[[1]], tb %>% pull(x), tb[["x"]]
- Data frame = named list of **same-length vectors**
- Iteration:
    - map() → return output
    - walk() → side effect

VI. Logical & Numeric Quirks

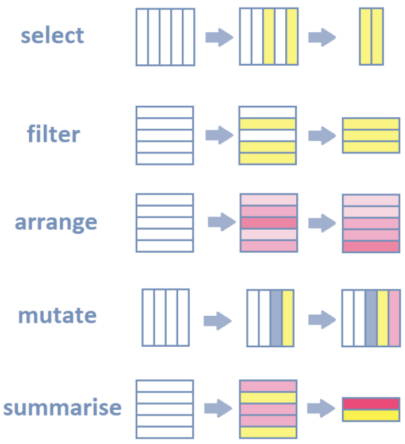
Expression	Result	Note
sqrt(2)^2 == 2	FALSE	Floating point precision
TRUE & NA	NA	
`TRUE	NA`	TRUE
sum(c(TRUE,TRUE,FALSE,TRUE))	3	

Recycling: short vector repeats automatically in arithmetic

Data Wrangling + Strings + Functions + Missing Data : Side 2 of the Sheet

Wrangling Essentials

- if_else(cond, yes, no) → vectorized condition
- parse_number() → extract numeric from messy strings
- Rounding: round(530.3,-2) ≈ (530.3 %/% 100) * 100



Factors:

- fct_relevel() → move levels to front
- fct_recode() → rename/combine
- fct_reorder() → reorder by numeric variable
- fct_infreq() → reorder by frequency

Dates:

- ISO8601 → yyyy-mm-dd hh:mm
- Convert strings → date-time class

Strings & Regex

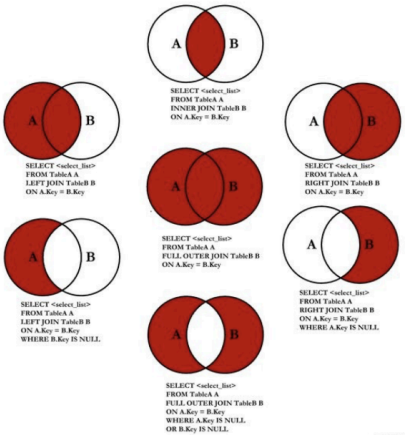
- Unicode / emoji: "Bird\n\tDuck\n\t\U1F6EA"
- Concatenate: str_c("Letter: ", letters, collapse='')
- Flatten: str_flatten(letters, ', ')
- Split: separate_wider_delim()
- Encoding: read_csv(..., locale=locale(encoding="Latin1"))

Regex:

- ? → optional
- + → 1+ repetitions
- Replace: str_replace_all("a/b/c","/","\\\\")

Relational Data

- Primary key → unique row ID
- Foreign key → links to primary key elsewhere
- Natural join → join on all common columns
- Non-equi join → inequality/cross joins



Missing Data

- Explicit → NA in dataset
- Implicit → missing rows
- tidyr::fill() → forward-fill explicit NA
- tidyr::complete() → generate missing rows
- NaN = Not a Number (similar to NA)
- Summary tips → use .drop=FALSE in group_by(), count(), scale_x_discrete()

Functions

- **Why:** readability, single update, reduce errors, reuse
- **Structure:** myfun <- function(arg) { body }
- Vector vs df functions → input & output differences
- Tidyverse → {{ }} around variables for tidy-eval
- := → programmatically create new variable

Example:

```
replicate_vector <- function(x,y) { rep(y, length.out=length(x)) }
```