7 Adv Data Wrangling P2

14 Strings

1. Understanding Strings In R
    - strings are text enclosed in quotes
    - use double quotes most of the time
    - escape special characters with a backslash
2. Combining And Splitting Strings
    - use paste() or str_c() to combine strings
    - use sep or collapse to control how strings join
    - use str_split() to separate strings into pieces
3. Measuring And Extracting Text
    - use str_length() to count characters
    - use str_sub() to extract or replace parts of a string
    - negative indices count from the end of the string
4. Detecting And Replacing Text
    - use str_detect() to see if a pattern exists
    - use str_replace() or str_replace_all() to substitute text
    - pattern matching is more powerful with regex
5. Working With String Vectors
    - stringr functions are vectorized
    - operations automatically apply to each element
    - results stay aligned with the original data

15 Regular Expressions

1. What Regular Expressions Are
    - regex is a language for describing text patterns
    - used for searching, detecting, and extracting text
    - essential for cleaning messy data
2. Literal Matching
    - plain characters match themselves
    - use \ to escape special characters
    - exact matches require no additional syntax
3. Character Classes And Quantifiers
    - use [] to match any of several characters
    - use \d, \s, and \w for digits, whitespace, and word characters
    - use +, *, and ? to control how many times a pattern appears
4. Anchors And Boundaries
    - ^ matches the start of a string
    - $ matches the end of a string

- \b matches a word boundary

5. Grouping And Alternation
    - parentheses create groups
    - use | to represent “or” patterns
    - grouping controls how matches are returned

19 Joins

1. Understanding Table Joins
    - joins combine two data frames based on keys
    - keys must identify how rows correspond
    - mismatched or missing keys affect results
2. Inner And Outer Joins
    - inner_join() keeps only matching rows
    - left_join() keeps all rows from the left table
    - right_join() keeps all rows from the right table
    - full_join() keeps all rows from both tables
3. Filtering Joins
    - semi_join() keeps rows in the left table that have matches in the right
    - anti_join() keeps rows in the left table that do not have matches
    - filtering joins do not add columns
4. Keys And Uniqueness
    - primary keys uniquely identify rows
    - if keys are not unique, joins may duplicate rows
    - always check for duplicates before joining
5. Matching Columns And Renaming
    - use by = to specify the join keys
    - R automatically matches keys with the same name
    - suffixes like .x and .y prevent conflicts when columns share names