

Kate Messitte – Exam I Summary Sheet

0. Data Structures in R

Vectors

- Homogeneous data
- `c()` to combine elements
- `seq()` for custom sequences

Lists

- Can hold heterogeneous objects
- Access:
 - `mylist[1]` → element
 - `mylist$element` → named element

Data Frames

- Create: `data.frame()`, `read.csv()`
- Access: `df$col`, `df[1,2]`

Tibbles

- Modern data frames: nicer printing, better handling of strings/factors

Categorical Variables

- Use `factor()` and `levels()`

1. Data Wrangling

- `select()`, `filter()`, `arrange()`, `mutate()`, `summarize()`, `group_by()`
- Pivot: `pivot_longer()`, `pivot_wider()`
- Joins: `left_join()`, `right_join()`

2. Data Visualization

Grammar of Graphics

- `ggplot(data) + aes(x, y) + geom_*()`

Geoms

- `geom_point()`, `geom_line()`, `geom_bar()`, `geom_histogram()`, etc

Facets

- `facet_wrap()`

Themes

- `theme_minimal()`, `theme_bw()`, etc

Labels

- `labs(title="...", x="...", y="...")`

3. Strings (from ThinkCSpy)

- Sequence of characters: `s[i]`, slicing
- Methods: `toupper()`, `tolower()`, `strsplit()`, `replace()`, `split()`
- Searching: `grep()`, `sub()`, `gsub()`, `grepl()`

4. Spatial Visualization

- **Spatial data handling**
 - `sf` package: `st_read()`, `st_write()`
 - Geometry types: points, lines, polygons
- **Basic plotting**
 - `plot()` for quick spatial plots
 - `ggplot2` + `geom_sf()` for layered spatial visualizations
- **Coordinate systems**
 - `st_transform()` to change CRS
- **Spatial operations**
 - `st_buffer()`, `st_intersection()`, `st_union()`
- **Mapping tips**
 - Color and size mappings: `aes(fill=..., size=...)`
 - Themes for maps: `theme_void()`, `coord_sf()`