SOC 4650/5650: Lecture 03 Functions

Christopher Prener, Ph.D.

February 5th, 2018

Packages

- `janitor`
- `naniar`
- `tidyverse`
  - `dplyr`
  - `magrittr`
  - `readr`

Reading `csv` Data

```
readr::read_csv(filePath
```

Verbs for Cleaning Data

Rename Variables

```
dplyr::rename(.data, newVar = oldVar)
```

Subset Data, Specific Observations

```
dplyr::filter(.data, expression)
```

Subset Data, Keep Specific Variables

```
dplyr::select(.data, varlist)
```

`varlist` items should be separated by commas

Subset Data, Drop Specific Variables

```
dplyr::select(.d, -varlist)
```

`varlist` items should be separated by commas with each item individually labeled with the dash drop symbol

Manipulate Variables

```
dplyr::mutate(.data, newVar = expression)
```

`ifelse` Outcomes

```
base::ifelse(expression, trueOutcome, falseOutcome)
```

Pipe Operator

Basic Syntax

`%>%` - "then"

Example with Assignment

```
mpg %>%
    select(manufacturer, model, cty, hwy) %>%
    rename(cityMpg = cty) %>%
    rename(hwyMpg = hwy) -> autoData
```

Remember that `dataFrame` names do not need to be included for many functions when included in piped code

Relational Operators

`<` - "less than"
`<=` - "less than or equal to"
`>` - "greater than"
`>=` - "greater than or equal to"
`==` - "exactly equal to"
`!=` - "not equal to"

Renaming Observations En Masse

```
janitor::clean_names(.data)
```

Renaming Observations En Masse

```
janitor::get_dupes(.data, varList)
```

Missing Data Analysis

Missing Variable Summary

```
naniar::miss_var_summary(.data)
```

Missing Case Summary

```
naniar::miss_case_summary(.data)
```