ART Forum R Cheat Sheet

Data Wrangling

Manipulating Observations

`filter()` observations
`arrange()` observations in ascending order
`arrange(desc())` observations in descending order
`bind_rows()` to stack observations

Manipulating Variables

`select()` variables
`mutate()` variables (recoded or new)
`transmute()` to mutate and select variables
`rename()` variables
`left_join()` matching observations from right df
`anti_join()` keeping observations that don't match
`inner_join()` keeping observations that match
`bind_cols()` to add variables
`gather()` variables into key-value variable pairs
`spread()` key-value variable pairs into variables
`separate()` a variable into two or more variables
`unite()` two or more variables into a single variable
`row_numbers()` to create an id for each observation

Data Summarization

Discrete Variables

`count()` discrete variables
`geom_col()` column plot
`geom_bar()` bar plot

Continuous Variables

`summarize()` continuous variables
`geom_histogram()` histogram
`geom_density()` density plot
`geom_point()` scatterplot
`geom_smooth(method = "lm")` regression line

Discrete and Continuous Variables

`group_by() %>% summarize()` grouped summary
`geom_boxplot()` boxplot
`geom_jitter()` jittered scatterplot
`facet_wrap()` and `facet_grid()` separate plots

Text Analysis

Data Wrangling

`unnest_tokens()` into a tidy text df
`get_sentiments()` from the specified dictionary
`bind_tf_idf()` to a df with counts by document
`read_html()` to scrape a webpage
`html_nodes()` extract data from a given XML node
`html_text()` extract text from a given XML node
`str_split()` split a string by a given regular expression
`str_view_all()` practice writing regular expressions

Topic Modeling

`cast_dtm()` from a tidy df with counts by document
`LDA()` run a topic model
`tidy(matrix = "beta")` topics from LDA output