Introduction to Cleaning Data in R

Andrew Hicks

2020-03-19

Data Science Process: Collect >> Clean >> Analyze >> Report

What we'll cover in this course:

1. Exploring raw data
2. Tidying data
3. Preparing data for analysis
4. Putting it all together

```
# View the first 6 rows of data
head(weather)

# View the last 6 rows of data
tail(weather)

# View a condensed summary of the data
str(weather)
```

Exploring RAW data:

1. **Understanding the structure of your data**

   ```
   # Load the lunch data

   lunch <- read.csv("datasets/lunch_clean.csv")
   ```

- **class() - Class of data object**

  ```
  # View its class

  class(lunch)

  > "data.frame"
  ```

- **dim() - Dimension of data**

```
# View its dimensions

dim(lunch)

> 46 7
```

- **names() - Column names**

```
# Look at the column names

names(lunch)

> "year" "avg_free" ... "perc_free_red"
```

- **str() - Preview of data with helpful details**

```
str(lunch) # provides dataframe row & columns, column's names, datatype and heading records

# Load dplyr

library(dplyr)
```

- **glimpse() - Better version of str() from dplyr**

```
# View structure of lunch, the dplyr way

glimpse(lunch)
```

- **summery() - Summary of data**

```
# View a summary

summary(lunch)
```

---

2. **Looking at your data**
3. **Visualize your data**