

Working with Data



Matthew Renze

DATA SCIENCE CONSULTANT

@matthewrenze www.matthewrenze.com



Overview



Working with Data

Demo



Working with Data

Import

Clean

Transform

Export



Importing Data

File-based data

Web-based data

Databases

Statistical data



Cleaning Data

Reshape data

Rename columns

Convert data types

Ensure proper encoding

Ensure internal consistency

Handle errors and outliers

Handle missing values



Cleaning Data

base

tidyr

reshape2

stringr

lubridate

validate



Transforming Data

Select columns

Filter rows

Group rows

Order rows

Merge tables



Transforming Data

base

plyr

dplyr

data.table

sqldf



Exporting Data

File-based data

Web-based data

Databases

Statistical data



Using dplyr

Select

Filter

Mutate

Summarize

Arrange



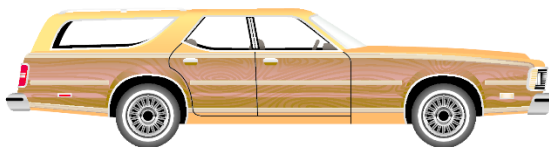
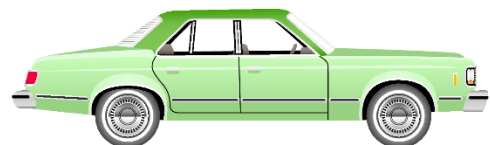
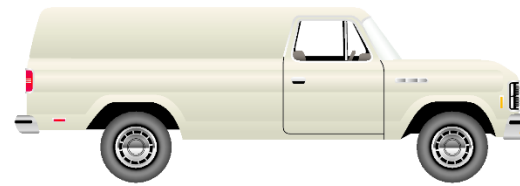
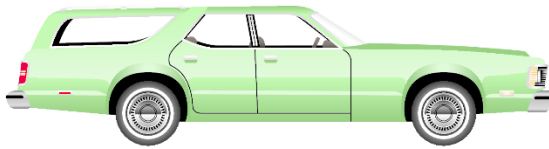
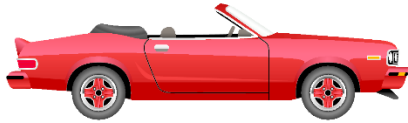
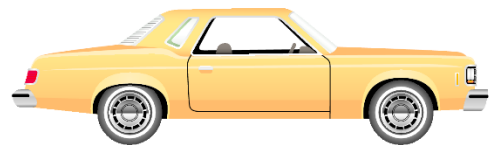
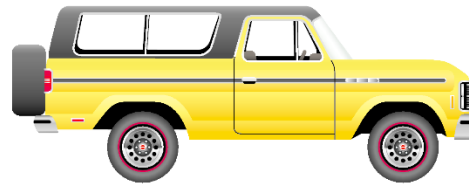
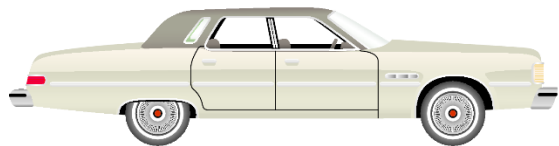
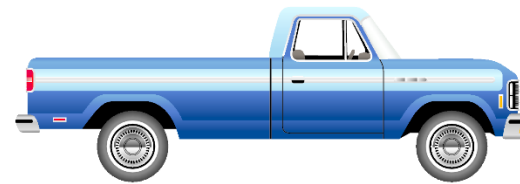
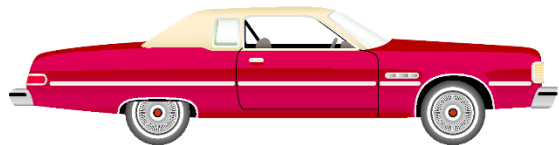
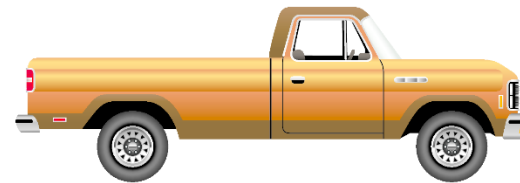
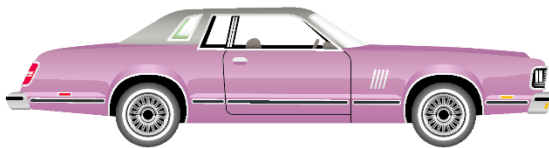
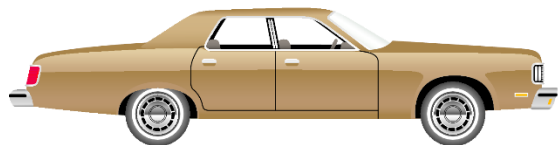
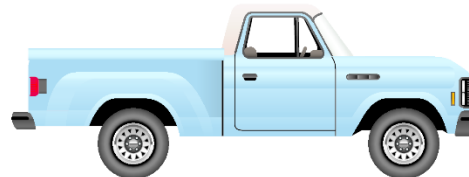
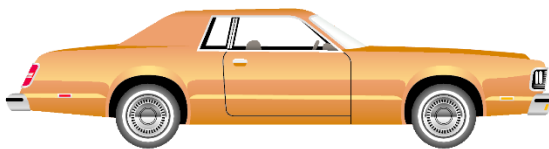
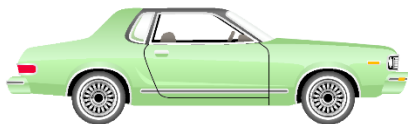
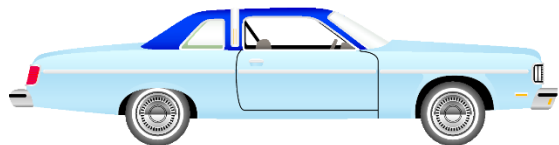
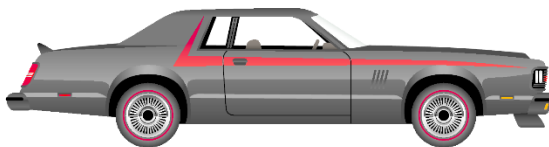
Advice for Working with Data

Often difficult

Time consuming

Record all steps





Motor Trend Data Set

Motor Trend Cars			
Name	Transmission	Cylinders	Fuel Economy
Mazda RX4	Manual	6	21.0
Mazda RX4 Wag	Manual	6	21.0
Datsun 710	Manual	4	22.8
Hornet 4 Drive	Automatic	6	21.4
Hornet Sportabout	Automatic	8	18.7
...



Demo Process



Inspect TAB file



Set the Working Directory



Load the Data



Peek at the Data



Load the dplyr package



Select Columns



Filter Rows



Mutate



Group



Summarize



Arrange



Convert to Data Frame



Chain Methods



Export Data



Inspect CSV File





Summary



Working with Data

Demo

