ABOUT DLOOKR PACKAGE

choonghyun ryu

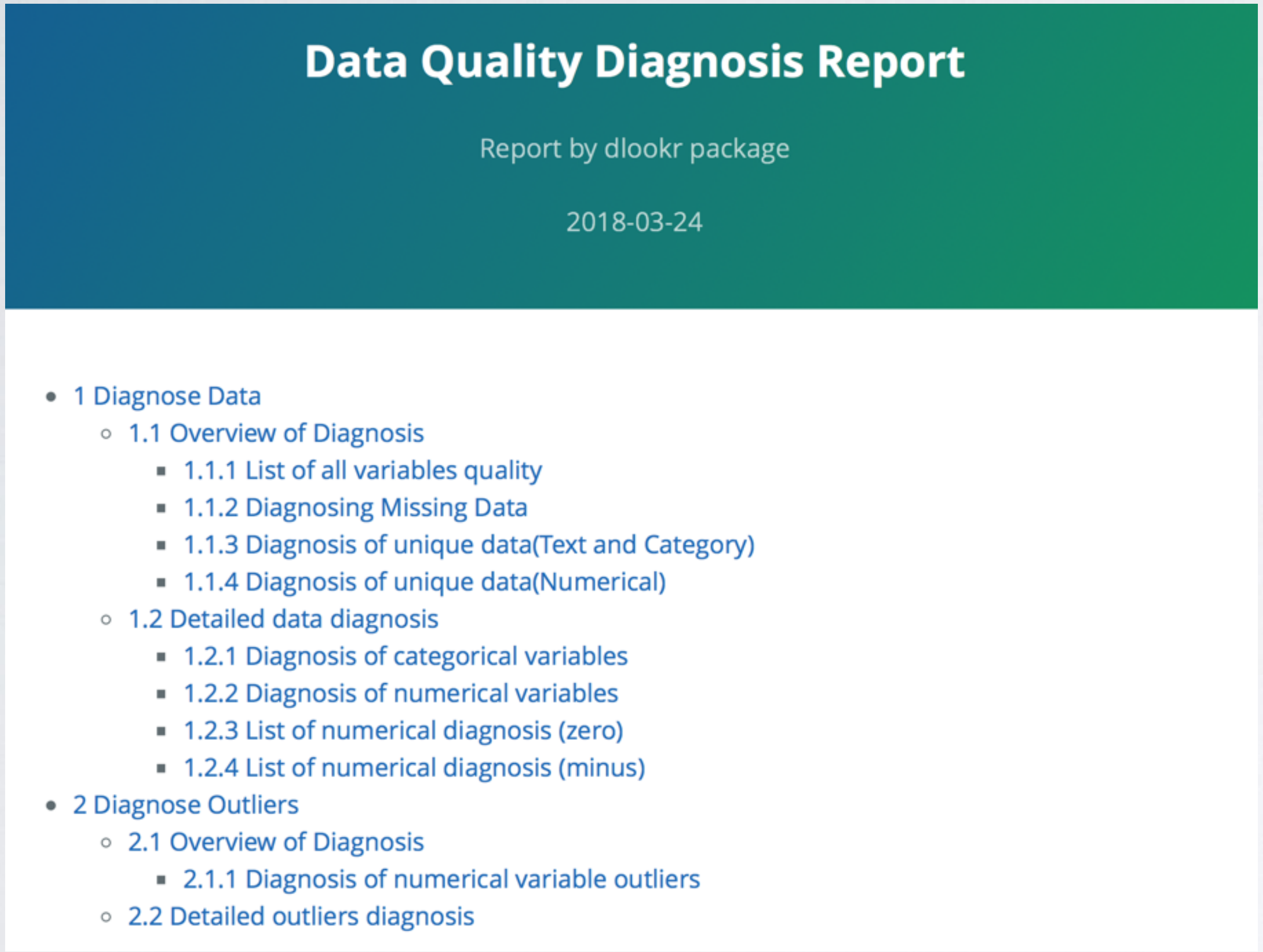
HISTORY

- who : choonghyun ryu
- when : after RStudio conference
- where : home
- what : R package
- why : motivation
- how : develop R package

ABOUT

- Package: dlookr
- Type: Package
- Title: Tools for Data Exploration & Data Transformation
- Version: 0.3.0
- Authors@R: c(person("Choonghyun", "Ryu", email = "choonghyun.ryu@gmail.com", role = c("aut", "cre")))
- Description: A collection of tools that support data diagnosis, exploration, and transformation.
- License: GPL-2

DATA DIAGNOSIS

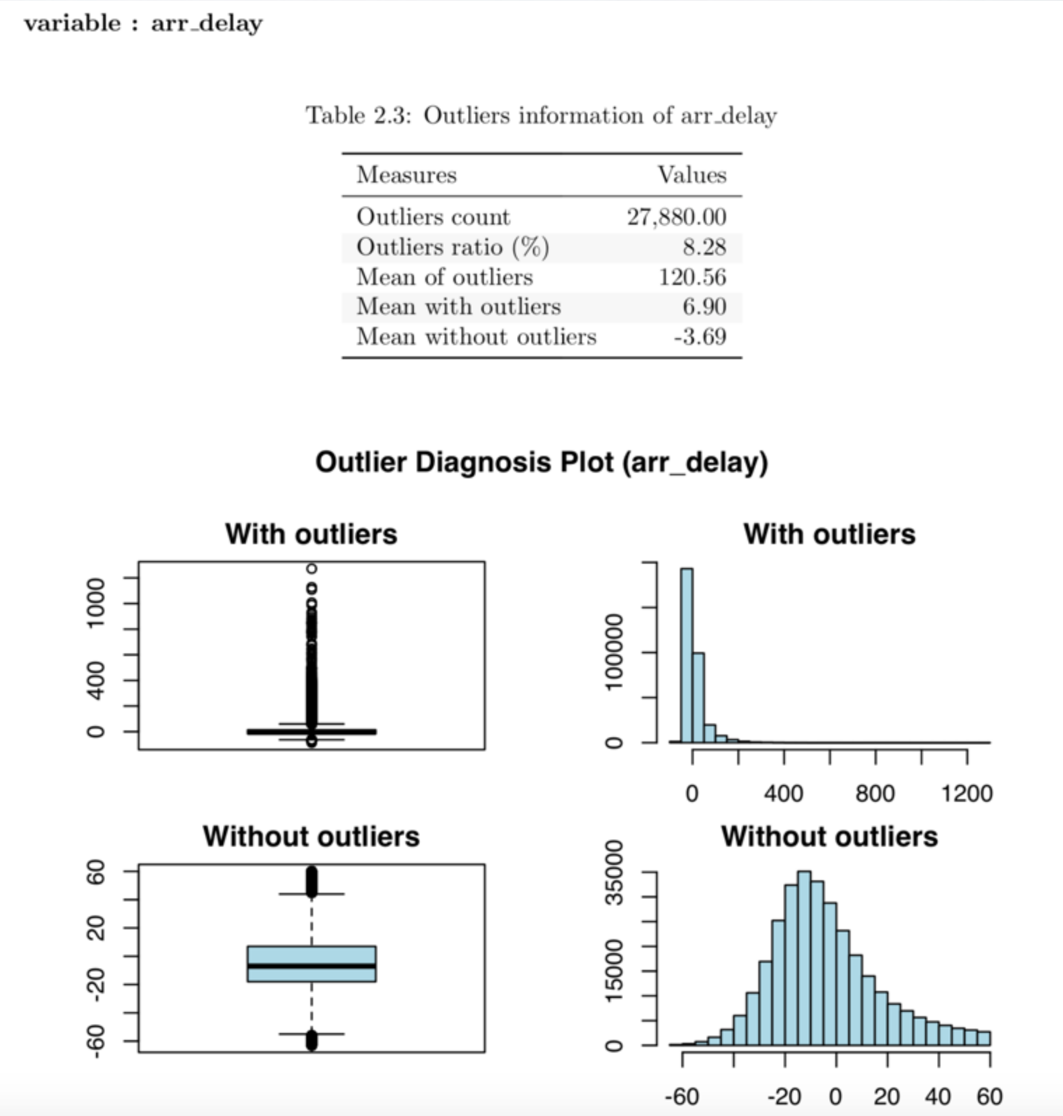


Data Quality Diagnosis Report

Report by dlookr package

2018-03-24

- 1 Diagnose Data
  - 1.1 Overview of Diagnosis
    - 1.1.1 List of all variables quality
    - 1.1.2 Diagnosing Missing Data
    - 1.1.3 Diagnosis of unique data(Text and Category)
    - 1.1.4 Diagnosis of unique data(Numerical)
  - 1.2 Detailed data diagnosis
    - 1.2.1 Diagnosis of categorical variables
    - 1.2.2 Diagnosis of numerical variables
    - 1.2.3 List of numerical diagnosis (zero)
    - 1.2.4 List of numerical diagnosis (minus)
- 2 Diagnose Outliers
  - 2.1 Overview of Diagnosis
    - 2.1.1 Diagnosis of numerical variable outliers
  - 2.2 Detailed outliers diagnosis



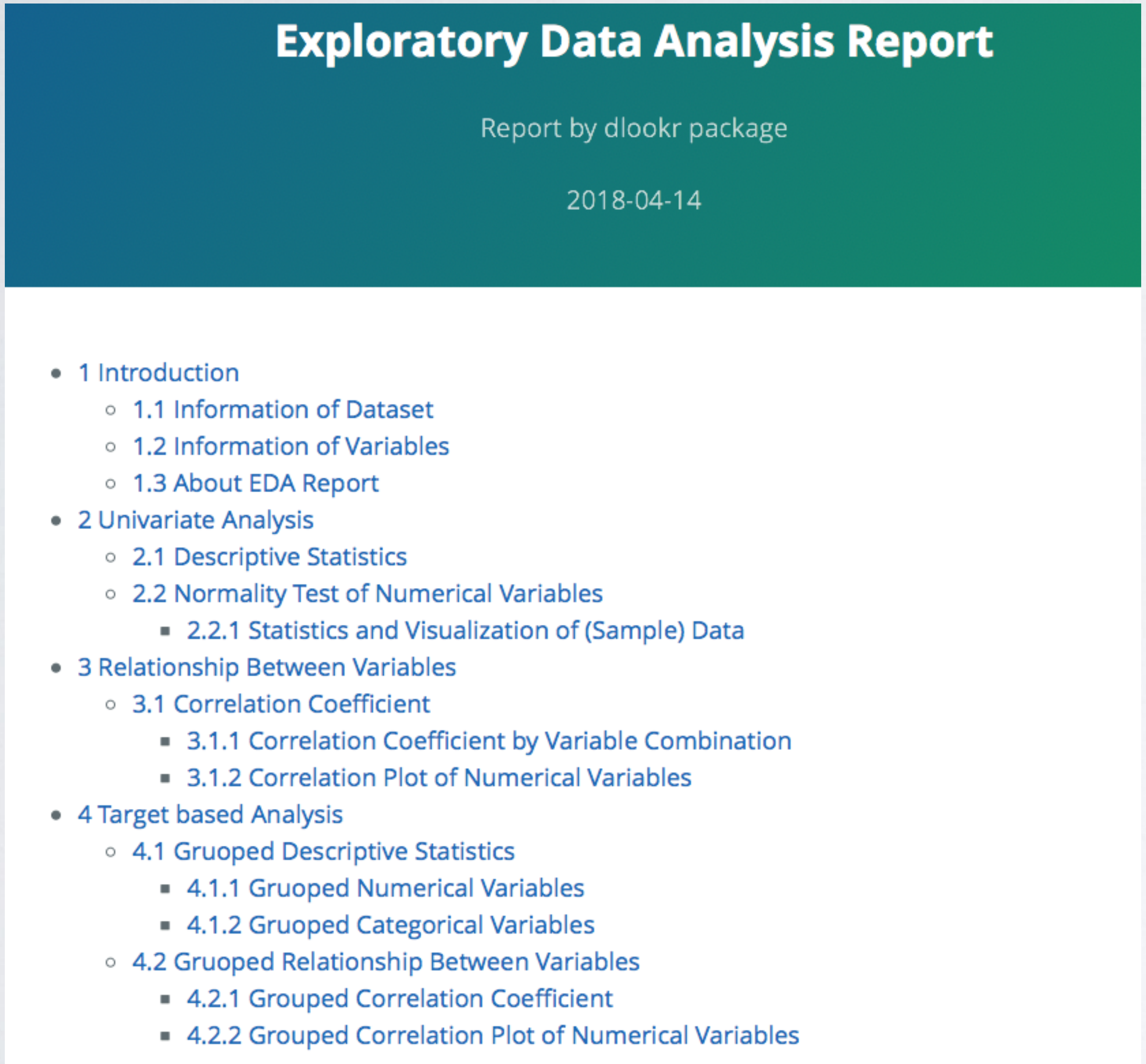
variable : arr_delay

Table 2.3: Outliers information of arr_delay

Measures	Values
Outliers count	27,880.00
Outliers ratio (%)	8.28
Mean of outliers	120.56
Mean with outliers	6.90
Mean without outliers	-3.69

Outlier Diagnosis Plot (arr_delay)

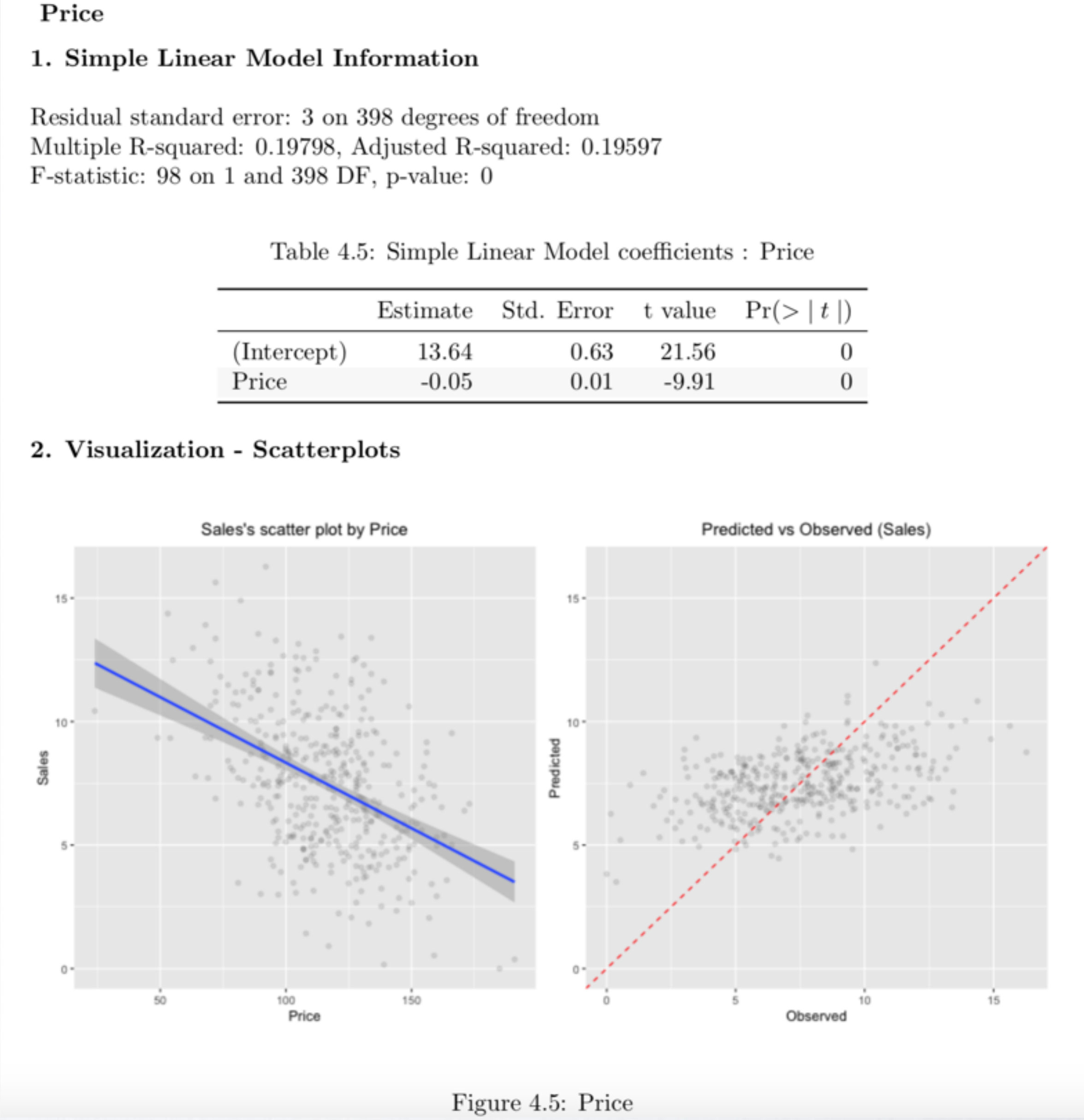
DATA EXPLORATION



Exploratory Data Analysis Report

Report by dlookr package

2018-04-14



Price

1. Simple Linear Model Information

Residual standard error: 3 on 398 degrees of freedom
Multiple R-squared: 0.19798, Adjusted R-squared: 0.19597
F-statistic: 98 on 1 and 398 DF, p-value: 0

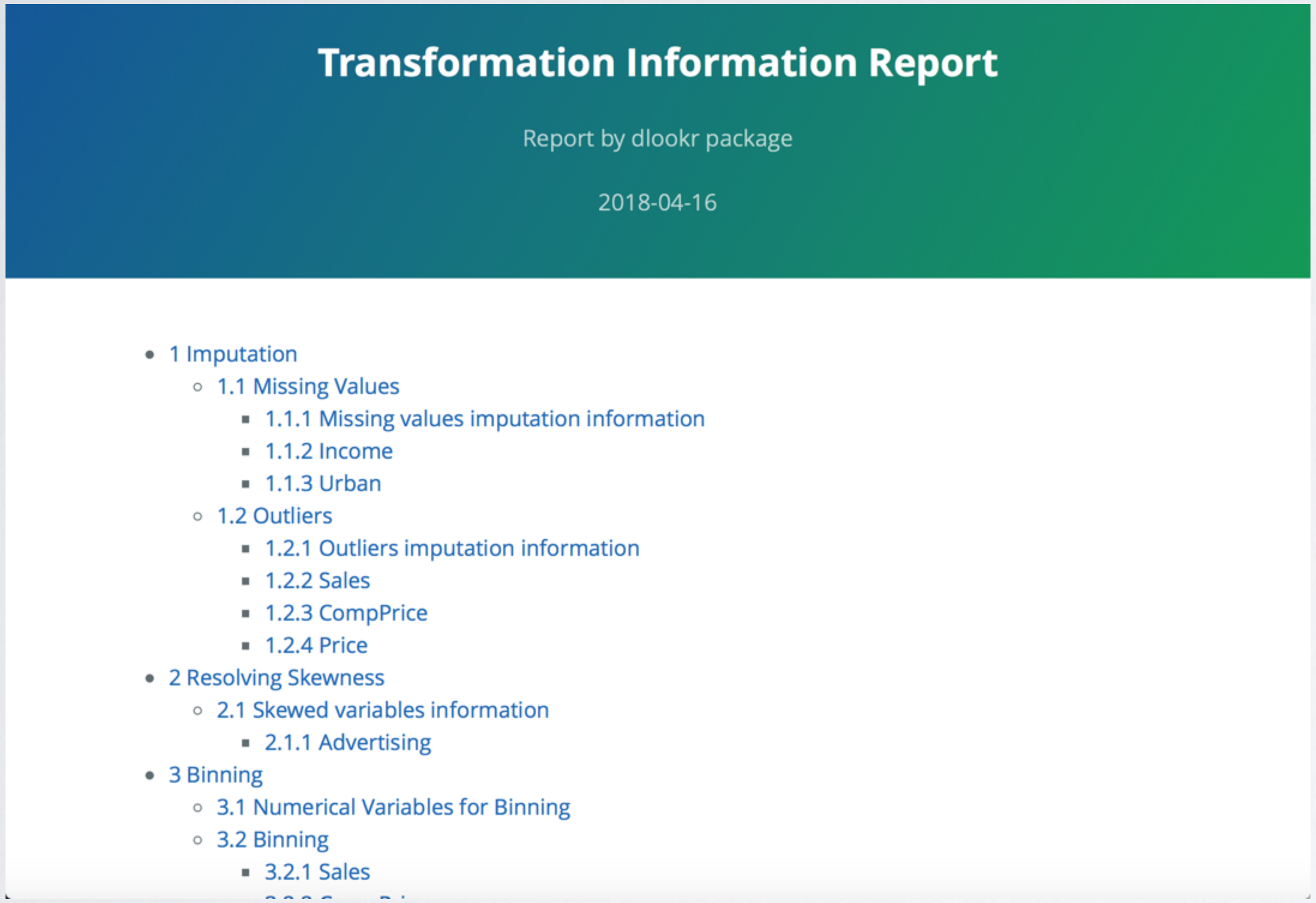
Table 4.5: Simple Linear Model coefficients : Price

	Estimate	Std. Error	t value	Pr(> | t |)
(Intercept)	13.64	0.63	21.56	0
Price	-0.05	0.01	-9.91	0

2. Visualization - Scatterplots

Figure 4.5: Price

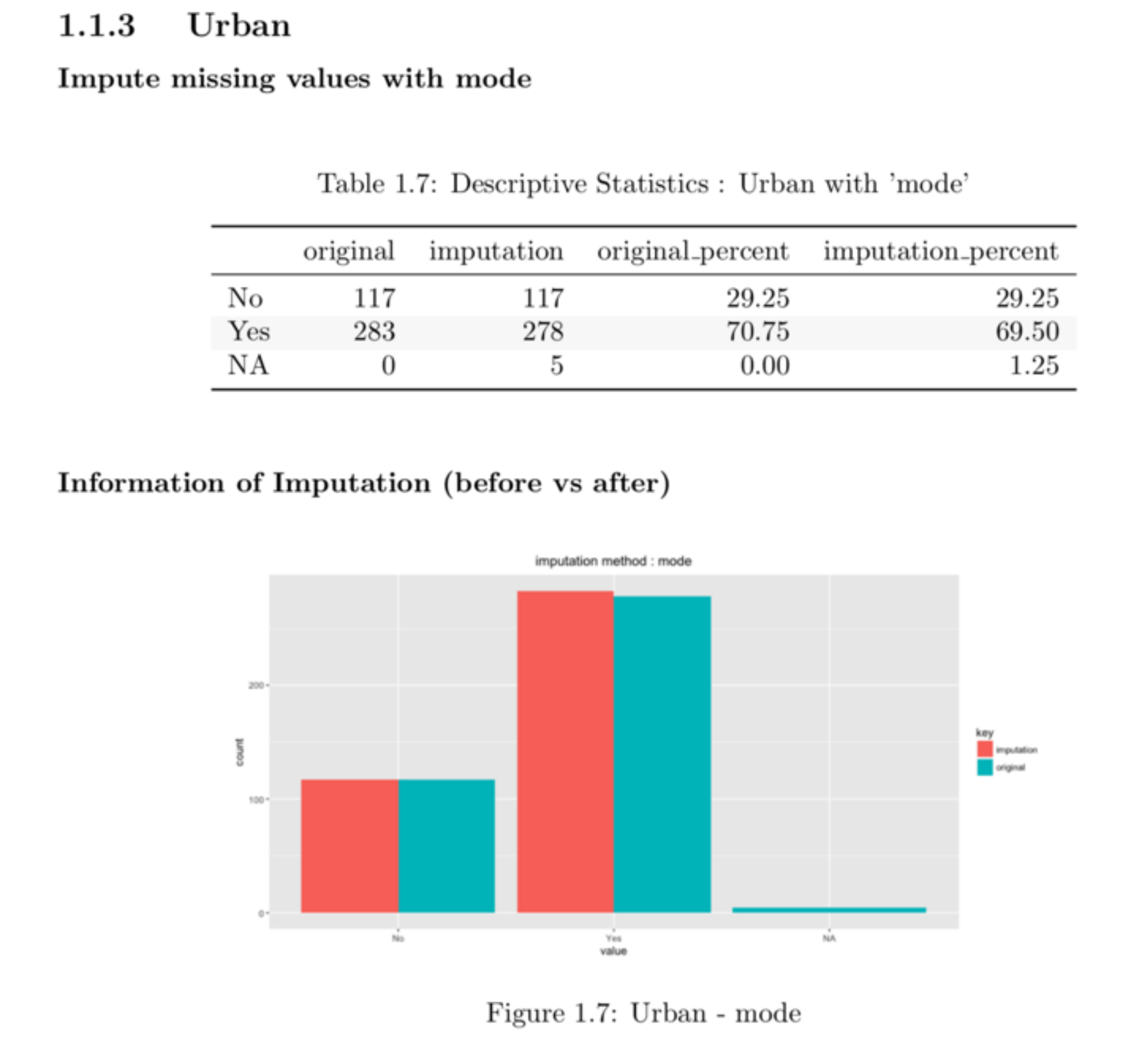
DATA TRANSFORMATION



Transformation Information Report

Report by dlookr package

2018-04-16



1.1.3 Urban

Impute missing values with mode

Table 1.7: Descriptive Statistics : Urban with 'mode'

	original	imputation	original_percent	imputation_percent
No	117	117	29.25	29.25
Yes	283	278	70.75	69.50
NA	0	5	0.00	1.25

Information of Imputation (before vs after)

Figure 1.7: Urban - mode

“pipe를 이용한 응용,
tidyverse와의 궁합.”

BUCKET LIST

- 작은 소망

  hexbin 스티커를 노트북에 붙이는 것