UCB Data Dictionary

Ben Croft

2022-09-07

Contents

Chapter 1

Introduction

This book is a prototype of a functional **Data Dictionary**.

As a proof of concept, the book serves as a demonstration of what University of Colorado Boulder may adopt as *open, collaborative, and living documentation* of data, data sets, and data applications. This proof of concept is version-controlled with Git, which enables version history, cross-team development, and rapid development. Additionally, this book can be hosted within Github Pages for public - or private - sharing among the campus system. Finally, this book can also be downloaded as a PDF, Ebook, or LaTeX file using the buttons at the top of this page.

1.1 Prerequisites

Since this data dictionary is written in Markdown, it can use anything that Pandoc's Markdown supports, e.g., a math equation $a^2 + b^2 = c^2$.

The **bookdown** package can be installed from CRAN or Github:

```
install.packages("bookdown")
# or the development version
# devtools::install_github("rstudio/bookdown")
```

Each Rmd file contains one and only one chapter, and a chapter is defined by the first-level heading `#`.

To compile this example to PDF, you need XeLaTeX. You are recommended to install TinyTeX (which includes XeLaTeX): https://yihui.name/tinytex/.

Chapter 2

Overview

The **Student Success Data Mart** provides the ability to review and manage analytics such as enrollment and registration metrics, count of student and faculty by registered courses, available courses for catalog building and schedule building, graduation rates and student career fulfillment of requirements, student academic standing, and faculty and student profiles.

Enrollment reports may include areas such as monitoring average class sizes, prerequisites not being met, grade distributions, and graduation eligibility.

Reports in this category can also include aggregated statistical reports with year-to-year comparisons in several subject areas.

With the Student Success Data Mart you can answer questions such as:

```
1) What are the enrollment metrics for this term?
2) Can I get the Student Retention and Graduation rates year over year by Cohort, Gender and Ethn
3) What is average GPA by Institution, Career and Program?
4) What is average time to graduate by Institution, Career and Program?
5) What classes are scheduled for this term?
6) Can I track and analyze the workload of the faculty?
7) What are the Class enrollment details?
8) What are the honors/awards details for the enrolled students?
```

2.1 Where does the Student Success Data Mart source its data from?

Student Records Data Mart is related to the Academic Program Activation and Management business process. This business process includes Course Catalog,

Class Scheduling, Student Career Term records and Enrollment business processes. These processes fulfill the institutions need to track course delivery, student participation through enrollments to those classes. The Academic Program activation and Management processes help manage class size, a student's enrollment in a class and track the resulting grades from the class.

Chapter 3

Academic Plan Summary Star

3.1 Description

This star contains student summary entry for a given academic plan and related academic program.

Each row of this table contains the most current information about an individual student and a particular academic plan. This fact table also contains all students that enrolled, matriculated, withdrew, or completed the academic career/program and plan in the institution.

The table is *not* keyed by term. `Institution`, `Career`, and `Program` are indirectly part of the key since they are derived from the `Academic Plan` dimension.

There is one row of data per student (person) academic plan surrogate ID and student career number.

Records are updated frequently, whenever the student program status changed. This may include status changes from activated, admitted, discontinued, or completed.

3.2 Helps Answer

This star may help answer the following:

- Number of enrollments by program or plan.
- Number of students who dropped out.

- Number of students in different program actions (statuses) and program action reasons by career, program, or plan.
- Number of students that completed the program/plan.
- Number of students that cancelled the program/plan.

3.3   Star Links

This star can be built from `F_ACADPLAN_SUMM` and connects to the following tables:

Table 3.1: A table generated by the longtable package.

matrix.input..ncol...1..byrow...T.
PS_D_ACAD_CAR
PS_D_ACAD_LOAD
PS_D_ACAD_ORG
PS_D_ACAD_PLAN
PS_D_ACAD_PROG
PS_D_ACAD_SPLAN
PS_D_ADMIT_TYPE
PS_D_AWD
PS_D_CAMPUS
PS_D_DAY
PS_D_DEG
PS_D_DEG_STAT
PS_D_INSTITUTION
PS_D_LOCATION
PS_D_PERSON
PS_D_PERSON_ADDR
PS_D_PERSON_ATTR
PS_D_PERSON_EMAIL
PS_D_PERSON_PHONE
PS_D_PROG_ACN
PS_D_PROG_ACN_RSN
PS_D_PROG_STAT
PS_D_STDNT_COHORT
PS_D_STDNT_GRP
PS_D_TERM
PS_F_ACADPLAN_SUMM
PS_R_ACAD_SPLAN
PS_R_AWD

PS_R_STDNT_GRP

Chapter 4

Term Enrollment Star

4.1 Description

This star contains information regarding term enrollments by student. The star provides term statistics and cumulative statistics by term, student, institution, and career. This star also provides measures such as units-in-progress, GPA, number of courses enrolled, etc.

Institution and Career are indirectly part of the key since they are deirved from the Term surrogate ID.

There is one row of data per term surrogate ID and student.

4.2 Helps Answer

This star may help answer the following:

- Top student academic standings.
- Nmber of units in progress vs. passed by institution, career, program, or plan.
- Average number of courses a student takes per term, by career, program, or plan.
- Number of units a student is enrolled in.
- Number of full time and part time students by term, institution, campus, primary program, etc.
- Number of students taking GPA units.
- Enrollment analysis by primary program and plan, term, institution, campus career, and program.

4.3 Star Links

This star can be built from `F_TERM_ENRLMT` and connects to the following tables:

Table 4.1: A table generated by the longtable package.

matrix.input..ncol...1..byrow...T.
PS_D_ACAD_LOAD
PS_D_ACAD_LVL
PS_D_ACAD_ORG
PS_D_ACAD_PLAN
PS_D_ACAD_PROG
PS_D_ACAD_STNDNG
PS_D_DAY
PS_D_PERSON
PS_D_PERSON_ADDR
PS_D_PERSON_ATTR
PS_D_PERSON_EMAIL
PS_D_PERSON_PHONE
PS_D_PROG_ACN
PS_D_PROG_ACN_RSN
PS_D_PROG_STAT
PS_D_RSDNCY
PS_D_STDNT_COHORT
PS_D_STDNT_GRP
PS_D_TERM
PS_D_YEAR
PS_R_PERSON_RSDNCY
PS_R_STDNT_GRP

Chapter 5

Applications

Some *significant* applications are demonstrated in this chapter.

5.1 Example one

5.2 Example two

Chapter 6

Final Words

We have finished a nice book.

Chapter 7

Introduction

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 7. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter 8.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```r
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure 7.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 7.1.

```r
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2022) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).
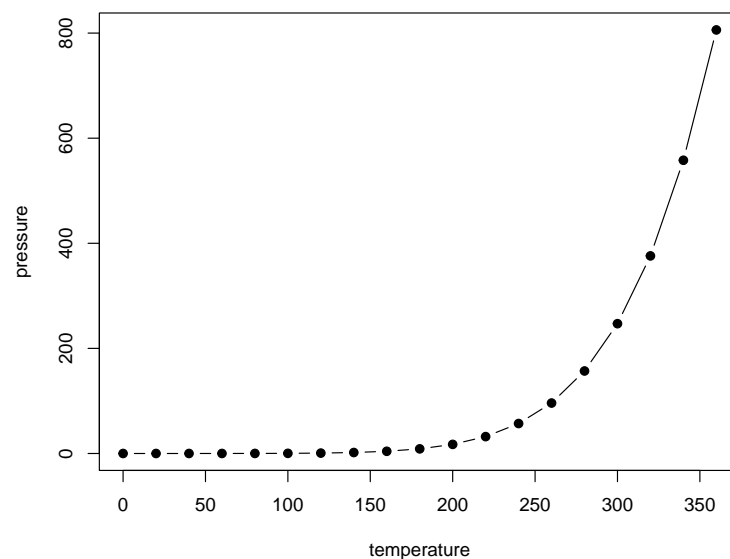
Figure 7.1: Here is a nice figure!

Table 7.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

Chapter 8

Methods

We describe our methods in this chapter.

Math can be added in body using usual syntax like this

8.1 math example

p is unknown but expected to be around 1/3. Standard error will be approximated

$$SE = \sqrt{(\frac{p(1-p)}{n})} \approx \sqrt{\frac{1/3(1-1/3)}{300}} = 0.027$$

You can also use math in footnotes like this[1].

We will approximate standard error to 0.027[2]

---

[1] where we mention $p = \frac{a}{b}$

[2] p is unknown but expected to be around 1/3. Standard error will be approximated

$$SE = \sqrt{(\frac{p(1-p)}{n})} \approx \sqrt{\frac{1/3(1-1/3)}{300}} = 0.027$$

Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2022). *bookdown: Authoring Books and Technical Documents with R Markdown*. https://github.com/rstudio/bookdown.