

Exploration

Nick Lauerman

January 22, 2018

Contents

Purpose	1
Libraries used	1
Data	1
Import the data	2
format the data	2
factors	2
Clean up reason	3
Date	3
Add new fields	3
Month and year	3
deliverable UID	3

Purpose

To explore metrics options based on data collected by Software Quality Assurance in a spreadsheet for document review and approval.

Libraries used

No libraries used yet

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
## The following object is masked from 'package:base':
##
##     date
```

Data

The data is presented in a CSV file (saved from an Excel workbook) with the following columns and characteristics

SQA SQA Preforming the review

Project ID Unique ID assigned to the project may come from NPV or ePAS

Application Application Name (limited selection)

Doc Version Version number of the document, also known as the revision number

Doc ID, if needed unique identifier to distinguish between documents for the same deliver type within a project

Approver or Disapproval? Approval status from the following list

* **A** Approved

* **A-FP** Approved on first pass

* **D** Disapproved

Date of Approval or Disapproval Date action was taken

Reason for Disapproval: Selected from the following list

- Inaccurate information
- Incorrect scope
- Insufficient testing
- Insufficient traceability
- N/A
- Not compliant with procedure
- Not following configuration management
- Not following document version control
- Not using template
- Other
- Requirement deficiency

Import the data

The data is read in and variables are assigned more programically useful names.

```
FPA.Raw <- read.csv(file = "./data/First Pass Acceptance.csv",
                    stringsAsFactors = FALSE,
                    col.names = c("sqa",
                                "project",
                                "application",
                                "deliverable",
                                "version",
                                "docID",
                                "status",
                                "date",
                                "reason",
                                "comments"))

FPA <- FPA.Raw
```

format the data

factors

Set the following data as factors:

*SQA

*Application

*Deliverable

*status

```
FPA$sqa <- as.factor(FPA$sqa)
FPA$application <- as.factor(FPA$application)
FPA$deliverable <- as.factor(FPA$deliverable)
FPA$status <- as.factor(FPA$status)
```

Clean up reason

convert blank fields in status to NA when status is converted to a factor

```
FPA$reason <- as.factor(FPA$reason)
levels(FPA$reason)
```

```
## [1] ""
## [2] "Inaccurate information"
## [3] "Incorrect scope"
## [4] "Not compliant with procedure"
## [5] "Not following document version control"
## [6] "Requirement deficiency"
```

```
levels(FPA$reason)[1] <- NA
levels(FPA$reason)
```

```
## [1] "Inaccurate information"
## [2] "Incorrect scope"
## [3] "Not compliant with procedure"
## [4] "Not following document version control"
## [5] "Requirement deficiency"
```

Date

```
FPA$date <- as.Date(FPA$date, format = "%d-%B-%y")
```

Add new fields

Month and year

From date create month and year fields

```
FPA$month <- month(FPA$date, label = TRUE)
FPA$year <- year(FPA$date)
```

deliverable UID

create an unique identification (UID) for each processed item so that each deliverable is uniquely identified accross the entire data set. to do this the UID will consist of the Project, deliverable, docID and version seperated with a dash

```
FPA$UID <- paste(FPA$project,
                 FPA$deliverable,
                 FPA$docID,
```

```
FPA$version,  
sep = "-"  
)
```