data_quality_assessment_malawi

ojgadabu

2/5/2019

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
#Define queries for extracting different data sets
#Start with unique patient identifiers for patients for a site
# Will use user stories to define the datasets required as below:
# Unique patient identifier user story
#
# As a patient enrolled in care
# I want to receive a unique patient identifier
# So that I can access services in the ART clinic
#
# Confirmations
# all patients with an internal patient id
# date that the patient was registered
# date that the encounter was created
# the identifier type of unique patient ide
# the patient identifier
#
# patient address current place of residence
#
# As a patient enrolled in care
# I need to have a current place of residence
# So that I can be contacted and retained in care
#
# Confirmations
# should have internal patient id
# should have identifier type
# should have the date the identifier was created
# should have the date the encounter was observed
# should have the current place of residence
# should be an active place of residence
#
# patient address phone number
#
# As a patient enrolled in care
# I need to have a phone number so that
# I may be reached when I need to remain in care
#
#Confirmations
# should have internal patient id
# should have identifier type
# should have the date the identifier was created
```

```
# should have the date the encounter was observed
# should have the contact phone number
# should be an active phone number
#
#As a patient enrolled in care
#I should have prescribed drugs at a patient visit
#So that I should receive a drug dispensation
#
#Confirmations
#Should have internal patient id
#Should have prescription encounter
#Should have name of prescribed drug
#Should have the date the prescription was done
#Should have the date the prescription was created
#
#
# As a patient enrolled in care
# I should receive a drug dispensation at my visit
# so that I have meds to take
#
# Confirmations
#Should have internal patient id
#Should have dispensation encounter
#Should have name of dispensed drug
#Should have the date the dispensation was done
#Should have the date the dispensation was created
#
#
# As a patient accessing care
# I should have a weight value at each visit
# So that my BMI may be monitored
#
# Confirmations
# Should have the internal patient id
# Should have the weights of the patient between study period
# should have the dates the weights were observed
# should have the dates the weights were created
#
#
# body mass index - height
# As a patient accessing care
# I should have my height measured at every visit
# So that my growth ratios may be calculated
#
# confirmations
# Should have the internal patient id
# Should have the weights of the patient between study period
# should have the dates the heights were observed
# should have the dates the heights were created
#
#
#
# viral load monitoring - viral load results
```

```
# As a patient accessing care
# I should have my viral load measured within 6 months of being enrolled in care
# So that I may gain viral load
# suppression
#
# confirmations
# Should have the internal patient id
# Should have the viral load results of the patient between study period
# should have the dates the viral load results were observed
# should have the dates the viral load results were created
#
#
#
#
```

Starting data quality assessment using the kahn framework with the data quality attributes of:

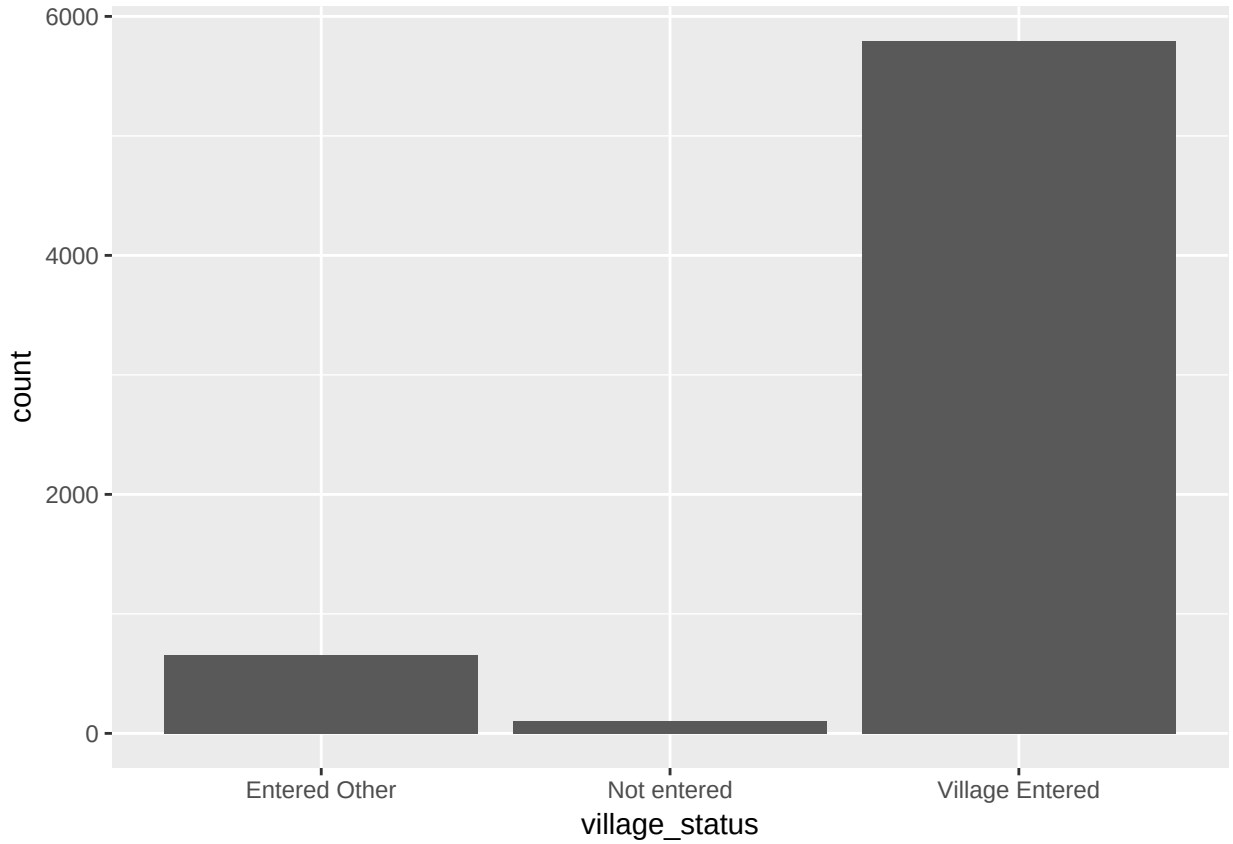
Completeness: Presence of data for an observation regardless of the structure or value of the data. For VL could be -100 or 700 or 200,000. Conformance: Adherence of data to a predefined format or structure of the data. This could be domain of values, for example, VL not having a value of -20 which would be outside the domain of defined values. Plausibility: This is the believability or truthfulness of observed data values, for example a patient with two viral load observations of 5000 and 700 within a space of a month.

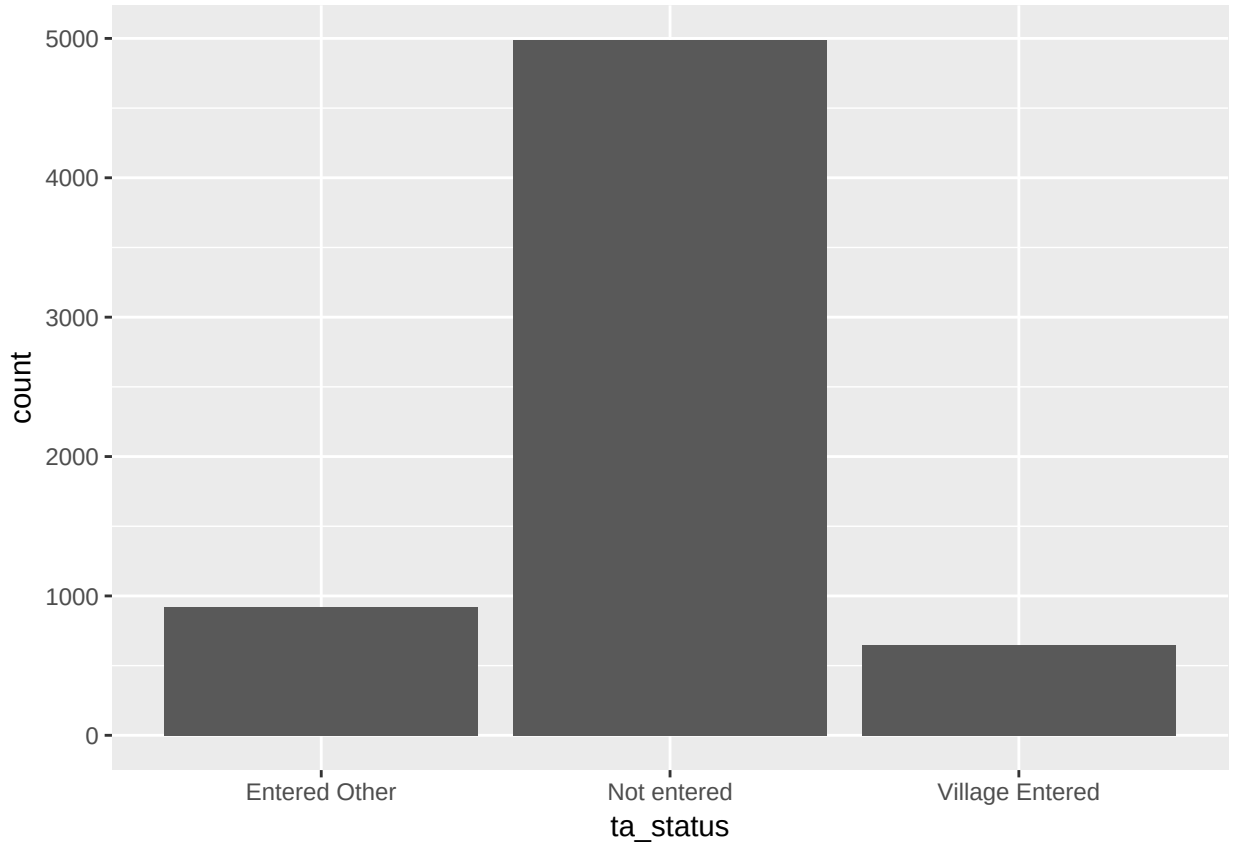
You can also embed plots, for example:

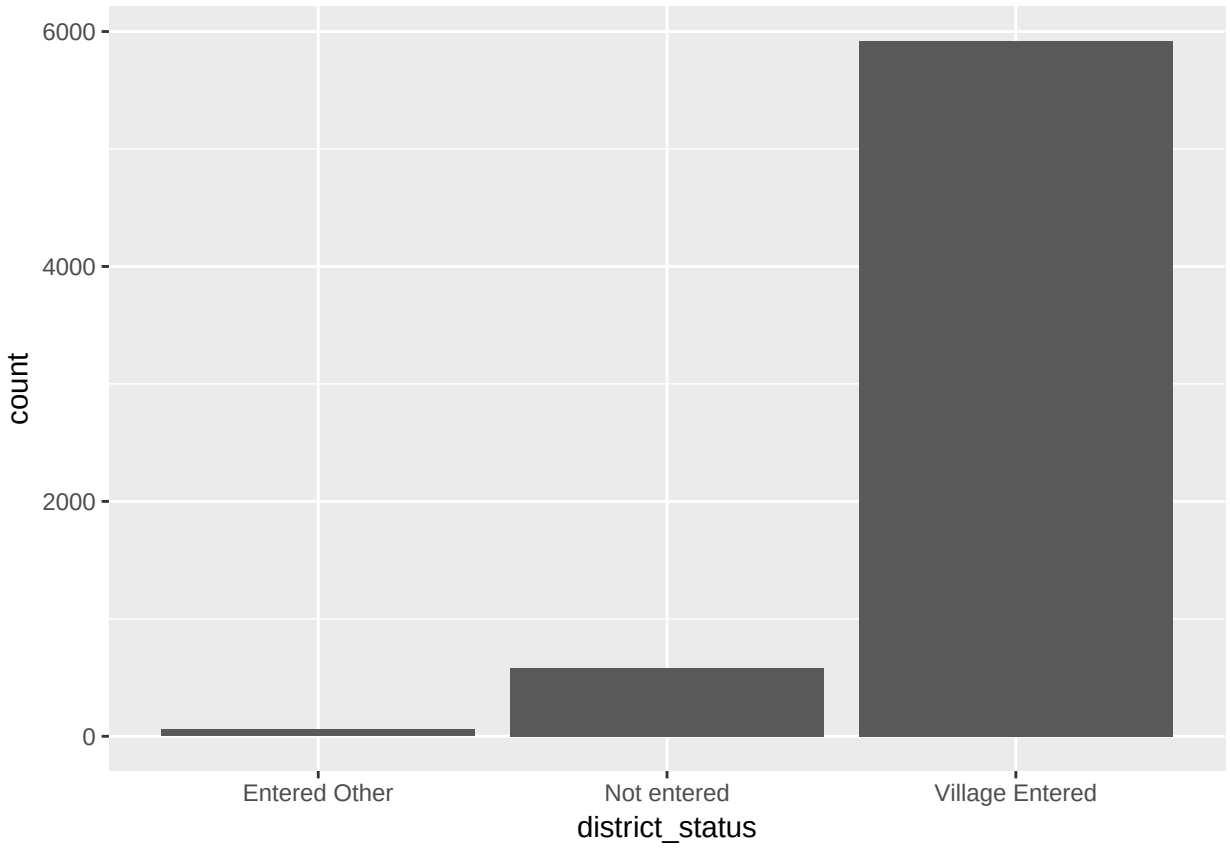
```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr
```

```
## Conflicts with tidy packages ---------------------------------------------
```

```
## filter(): dplyr, stats
## lag():    dplyr, stats
```


6000
4000
2000
0
count
Entered Other
Not entered
Village Entered
village_status


5000
4000
3000
2000
1000
0
count
Entered Other
Not entered
Village Entered
ta_status



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.