data_quality_assessment_malawi

ojgadabu

2/5/2019

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.

Results

*Completeness Table

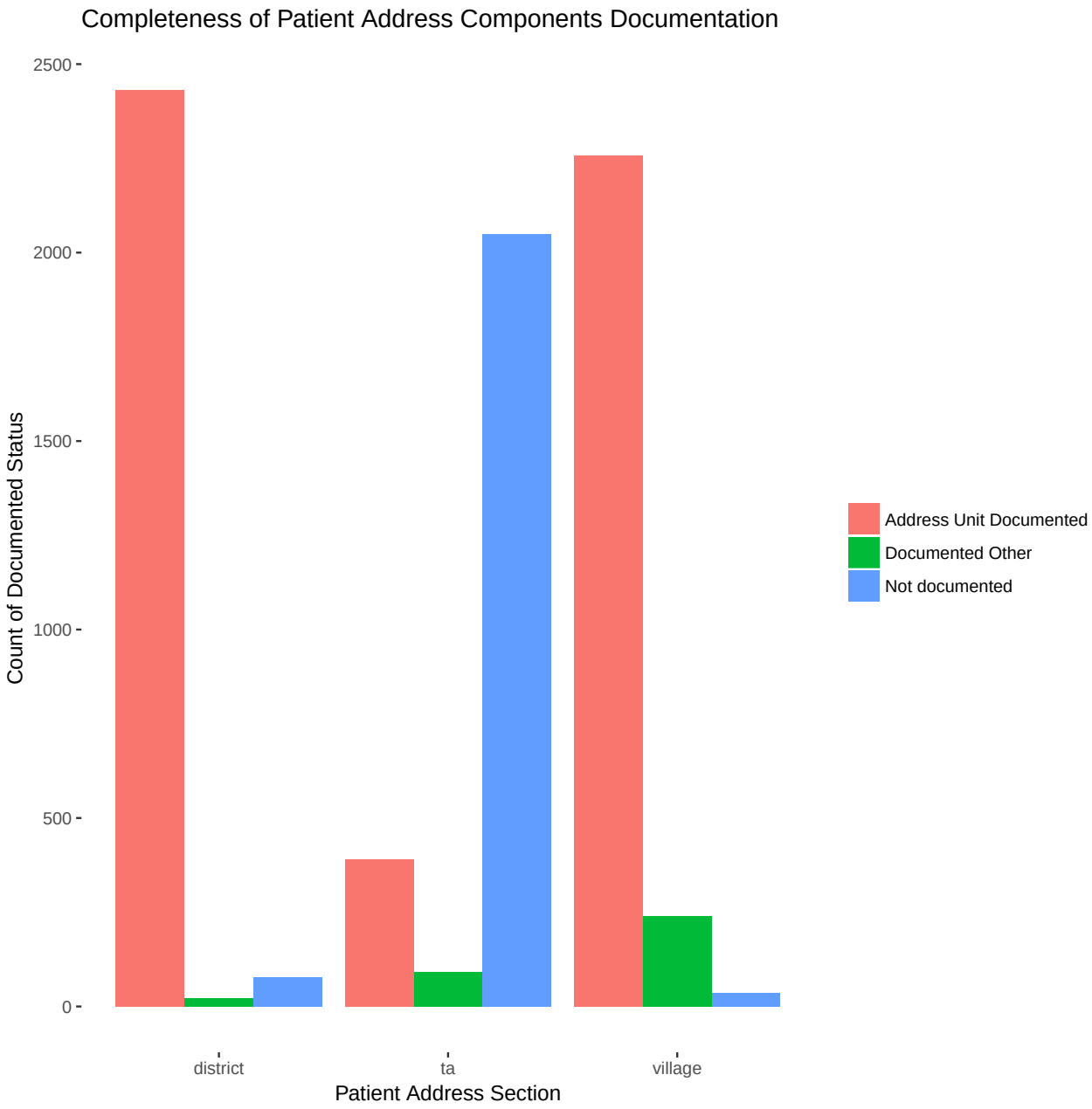
*Graph checking the completeness of address components. A complete address should have all components.

```
kable(completeness_table, caption = "Completeness of Documented Patient Records at a Selected ART Cli
```

Table 1: Completeness of Documented Patient Records at a Selected ART Clinic

Variable Name	Numerator Count	Denominator Count	Completeness Proportion
Current Patient Address	483	2532	19.08
Patient Mobile Phone Number	2410	2532	95.18
Weights of Patient	36309	36309	100.00
HIV Viral Load	14792	14795	99.98
Prescription Made	112505	120483	93.38
Drug Adherence	0	0	0.00

```
ggplot(cd) + geom_bar(mapping = aes(x = address_section, fill = address_value), position = "dodge") +
```



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