

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.

```
table1(~gender + age|complete_address, data = id_data_table_gech, output = "markdown")
```

```
## [1] "<table class=\"Rtable1\">\n<thead>\n<tr>\n<th class='rowlabel firstrow lastrow'></th>\n<th clas
```

```
kable(b, caption = "Table 1: Gender and Age by Health Facility Types of All Study Records")
```

Table 1: Table 1: Gender and Age by Health Facility Types of All Study Records

	x
District(n=37863)	
Health Center(n=38487)	
Tertiary(n=89333)	
Overall(n=165683)	
gender	
F	
24582 (64.9%)	
24649 (64.0%)	
52652 (58.9%)	
101883 (61.5%)	
Female	
1 (0.0%)	
0 (0%)	

3 (0.0%)
 4 (0.0%)
 M
 13280 (35.1%)
 13837 (36.0%)
 36675 (41.1%)
 63792 (38.5%)
 Male
 0 (0%)
 1 (0.0%)
 3 (0.0%)
 4 (0.0%)
 age
 Mean (SD)
 34.8 (10.6)
 34.0 (10.0)
 36.1 (10.2)
 35.3 (10.3)
 Median [Min, Max]
 33.0 [15.0, 87.0]
 33.0 [15.0, 89.0]
 35.0 [15.0, 99.0]
 34.0 [15.0, 99.0]
 |

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

Table 2: Completeness of Documented Patient Records at a Selected ART Clinic

Variable Name	Numerator Count	Denominator Count	Completeness Proportion
Current Patient Address	2254	29078	7.75
Patient Mobile Phone Number	12396	29078	42.63
Weights of Patient	261339	262570	99.53
HIV Viral Load	887	1095	81.00
Prescription Made	1038086	1048575	99.00
Drug Adherence	640331	640719	99.94

```
kable(all_healthfacility_completeness_table, caption = "Completeness of Documented Patient Records at Selected ART Clinics")
```

Table 3: Completeness of Documented Patient Records at Selected ART Clinics

Variable Name	Numerator Count	Denominator Count	Completeness Proportion
Current Patient Address	8022	165683	4.84
Patient Mobile Phone Number	85012	165683	51.31
Weights of Patient	1529898	1558136	98.19
HIV Viral Load	57435	64360	89.24
Prescriptions Made	0	0	0.00
Drug Adherence	2553881	2559997	99.76

```
kable(a, caption = "Gender, Age by Completeness of Physical Address of All Health Facilities")
```

Table 4: Gender, Age by Completeness of Physical Address of All Health Facilities

	—
	x
	—
Complete(n=2254)	
Incomplete(n=26824)	
Overall(n=29078)	
gender	
F	
1367 (60.6%)	
15361 (57.3%)	
16728 (57.5%)	
M	
887 (39.4%)	
11463 (42.7%)	
12350 (42.5%)	
age	
Mean (SD)	
35.2 (10.2)	
35.7 (10.1)	
35.6 (10.1)	
Median [Min, Max]	
34.0 [15.0, 78.0]	
34.0 [15.0, 96.0]	
34.0 [15.0, 96.0]	

```
kable(c, caption = "Gender, Age, Health Facility Type Records by Completeness of Physical Address")
```

Table 5: Gender, Age, Health Facility Type Records by Completeness of Physical Address

—
x
—

```
|
Complete(n=8022)
Incomplete(n=157243)
Overall(n=165683)
gender
F
4883 (60.9%)
96743 (61.5%)
101883 (61.5%)
Female
3 (0.0%)
1 (0.0%)
4 (0.0%)
M
3134 (39.1%)
60497 (38.5%)
63792 (38.5%)
Male
2 (0.0%)
2 (0.0%)
4 (0.0%)
age
Mean (SD)
34.3 (10.4)
35.4 (10.3)
35.3 (10.3)
Median [Min, Max]
33.0 [15.0, 79.0]
34.0 [15.0, 99.0]
```

34.0 [15.0, 99.0]

health_facility_type

District

2903 (36.2%)

34588 (22.0%)

37863 (22.9%)

Health Center

1465 (18.3%)

36998 (23.5%)

38487 (23.2%)

Tertiary

3654 (45.5%)

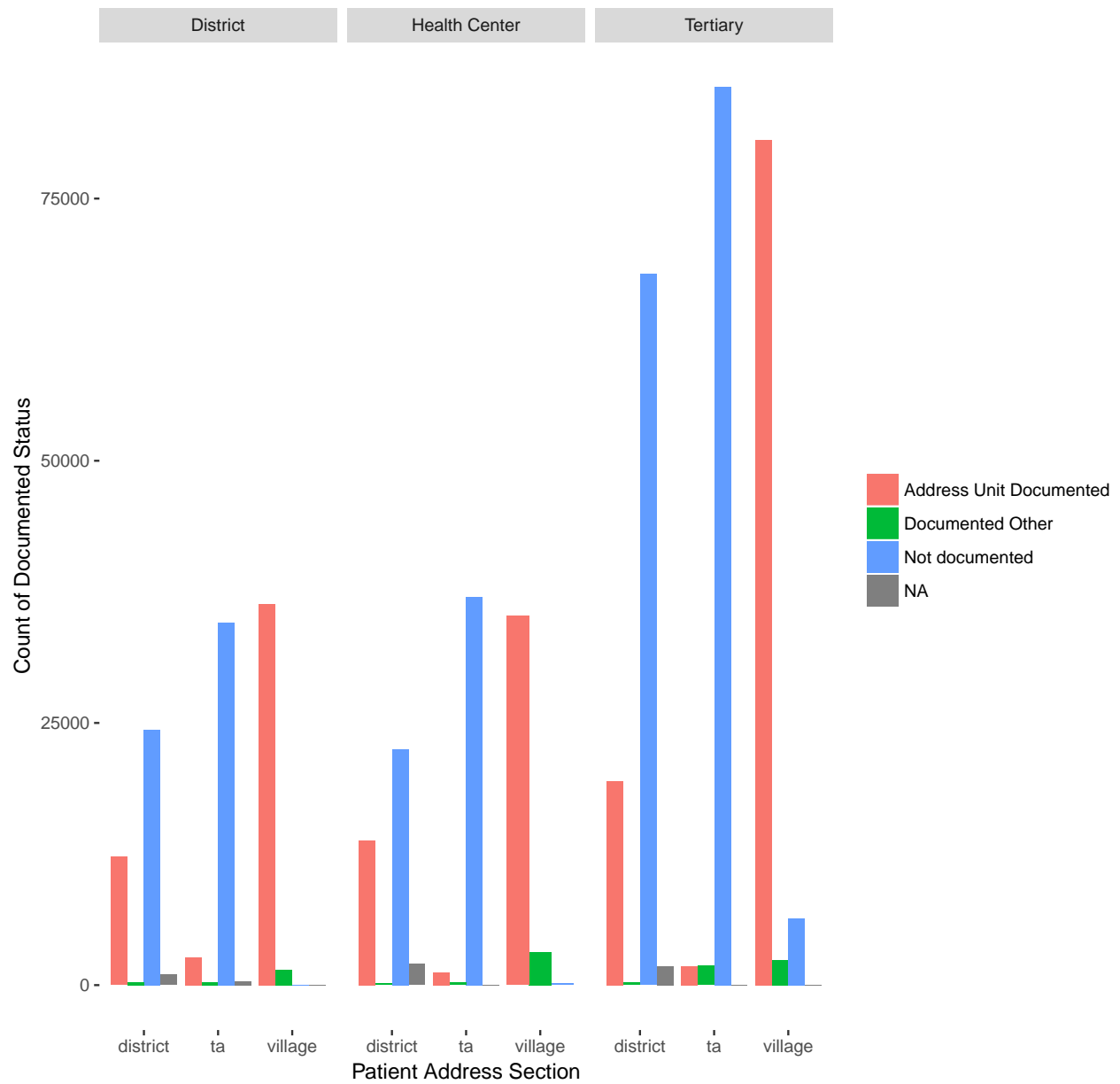
85657 (54.5%)

89333 (53.9%)

|

```
ggplot(cd) + geom_bar(mapping = aes(x = address_section, fill = address_value), position = "dodge") + 1
```

Completeness of Patient Address Components Documentation



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