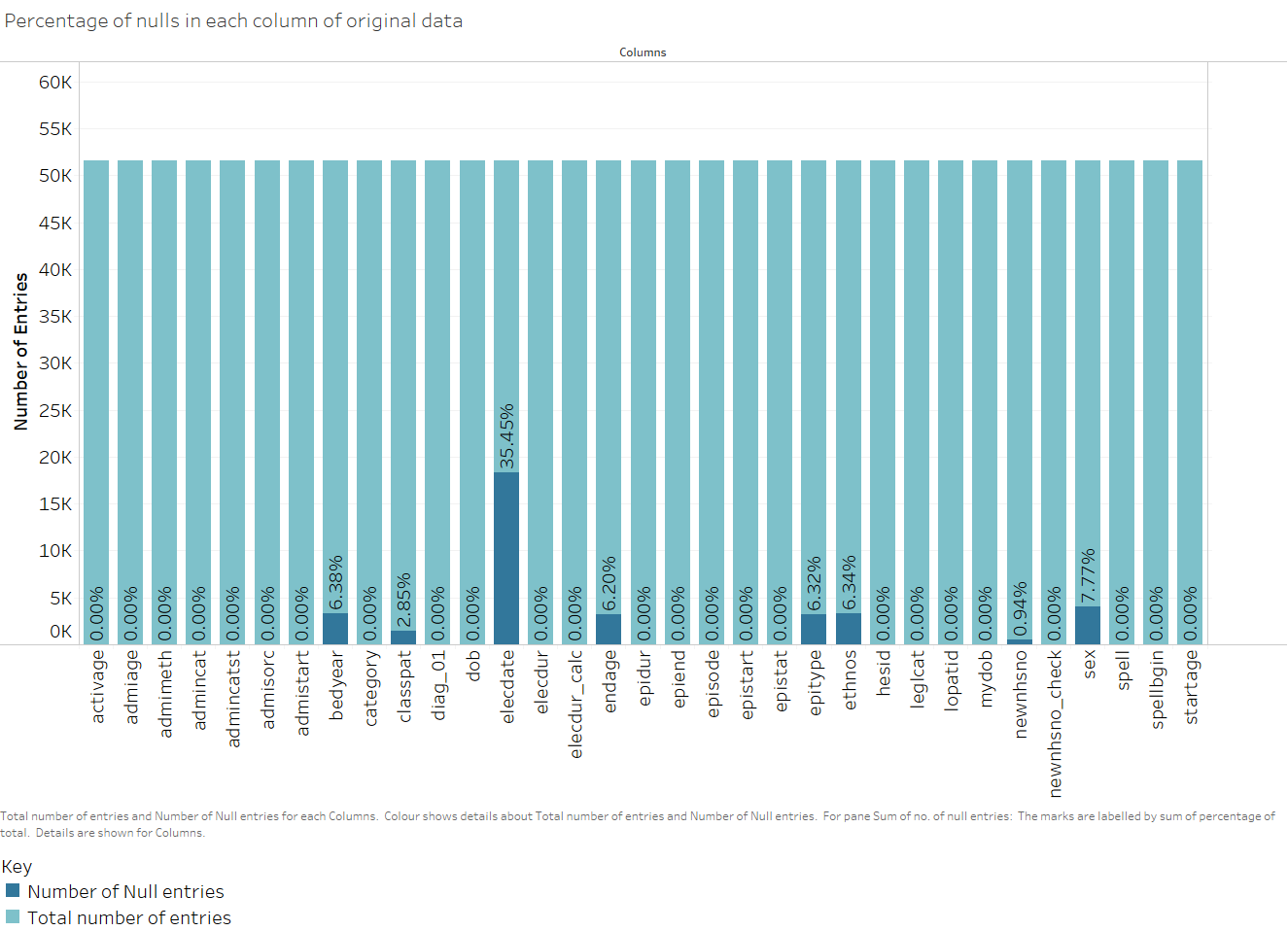
**Data Quality Visualisations**

Percentage of nulls per column in original data – the total data contains 2.189 of nulls



Overall:

Summary: This graph shows the number of null values in the graph across all columns

DQ Dimension: **Completeness**. % Null values can be used as a metric to indicate completeness with a lower percentage indicating better completeness.

# Sex

# 

Summary: The chart above indicates the distribution of patient gender for all entries (not for each patient)

Data Quality Dimension: **Completeness**: % Not known Patients (8%) is an indicator of completeness,

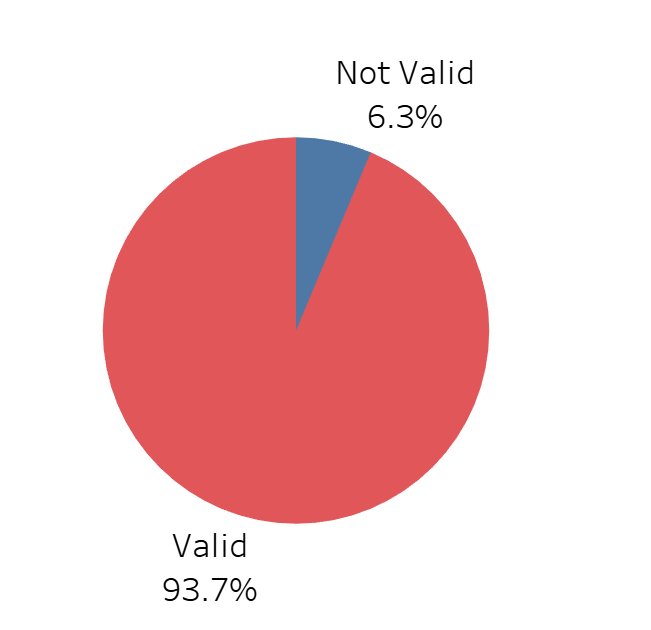
ElecDate (Date at which It was decided that patient would be admitted)

# 

Summary:The pie chart above indicates the type of entry for date of birth, either: ‘Valid’, if the date was valid; ‘Null’, if the date was 01/01/1800 or; ‘Invalid’ is the date was 01/01/1801. This classification is from the Data Dictionary

Dimension: **Completeness:**  34% of the dates were null indicated 66% completeness for this field.

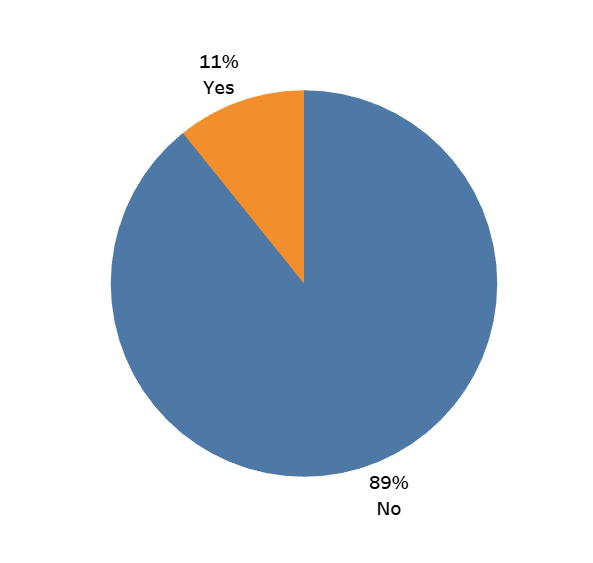
Epitype



Summary: The chart above shows the entries in epitype which did not match the valid entries defined in the Data Dictionary

Dimension: **Validity,** the chart above indicates 93.7% Validity.

Start Age

Is Start Age Correct?

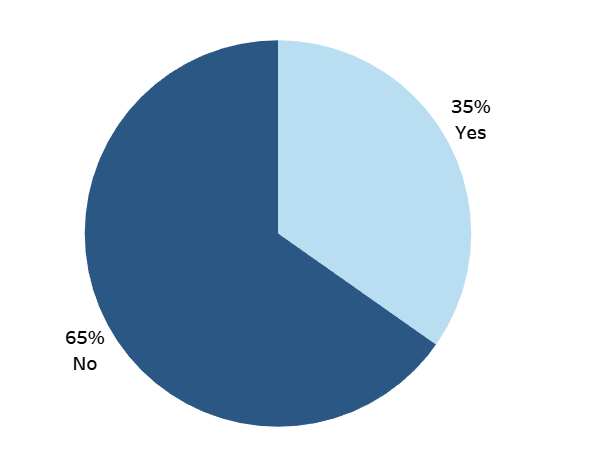


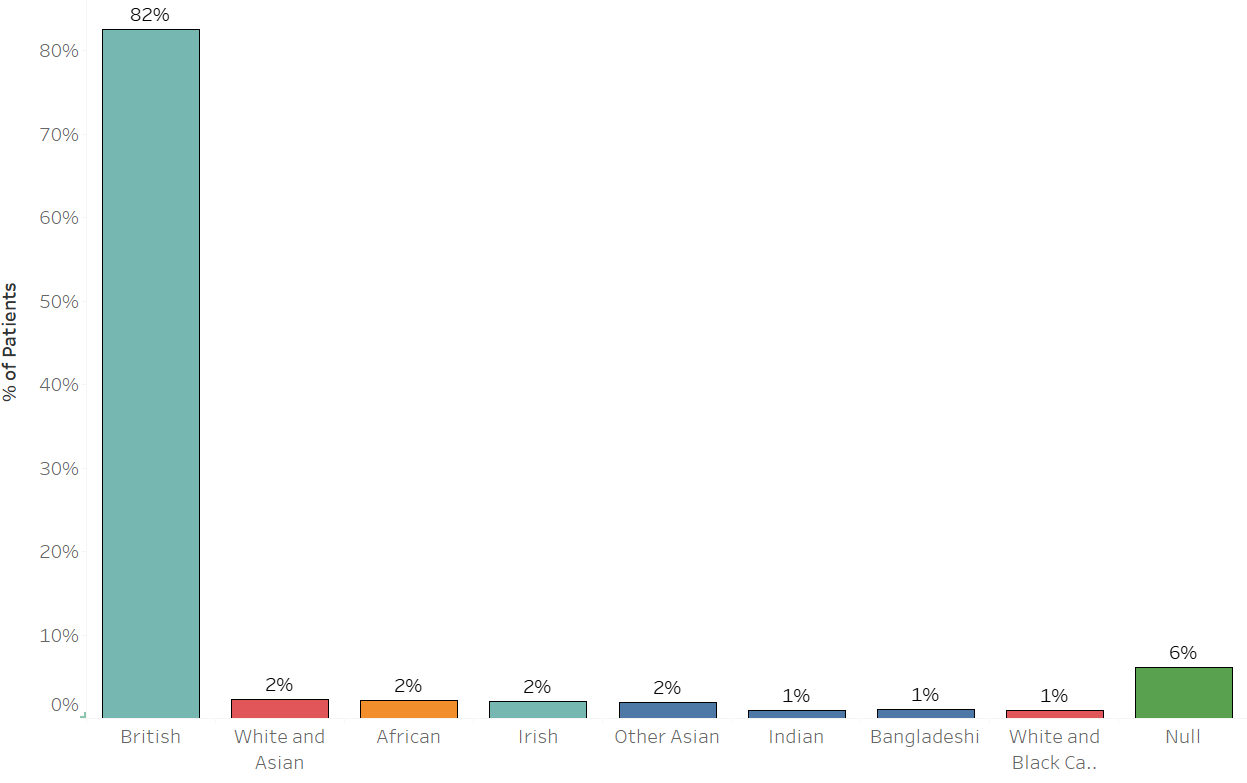
Figure :Correct Start Age Entries as % of Total Entries

Figure : Correct Start Age Entries as % of Distinct Patients

# Summary: The graphs above show on how many instances startdate is correct through comparison with the date of birth and start of episode. This has been calculated as a percentage of correct entries and also as a percentage of patients with correct entries.

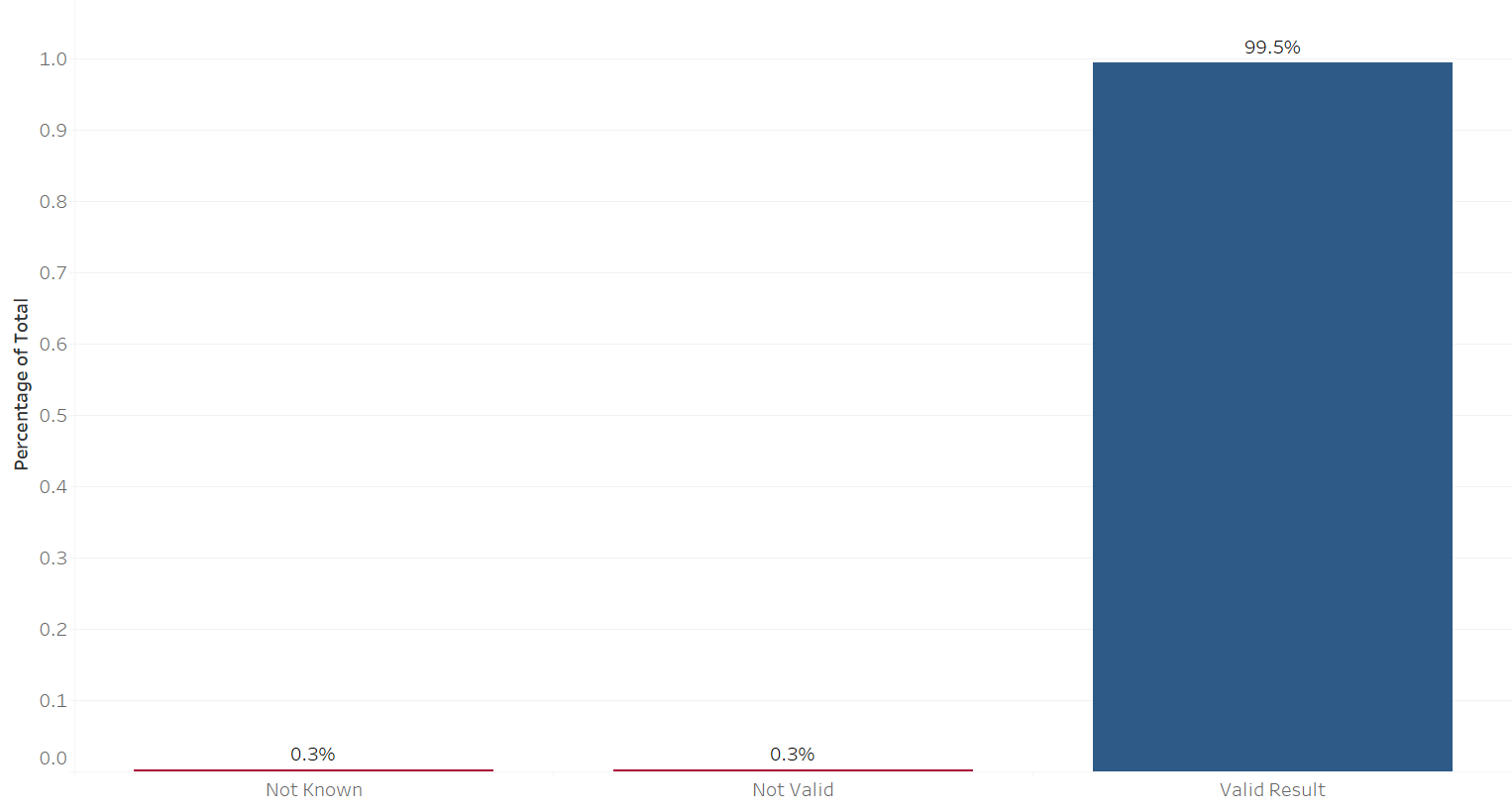
DQ Dimension: **Accuracy**. Using % of correct start dates as our measure of accuracy, we could say that this column has 11% accuracy, which is very low. In a normalised data set this category would be irrelevant as it could be calculated from date of birth.

Ethnos



Summary: The graph above shows the % of patients of different ethnicities. It should be noted that this counts individual patients and not the number of episodes of patients of different ethnicities.

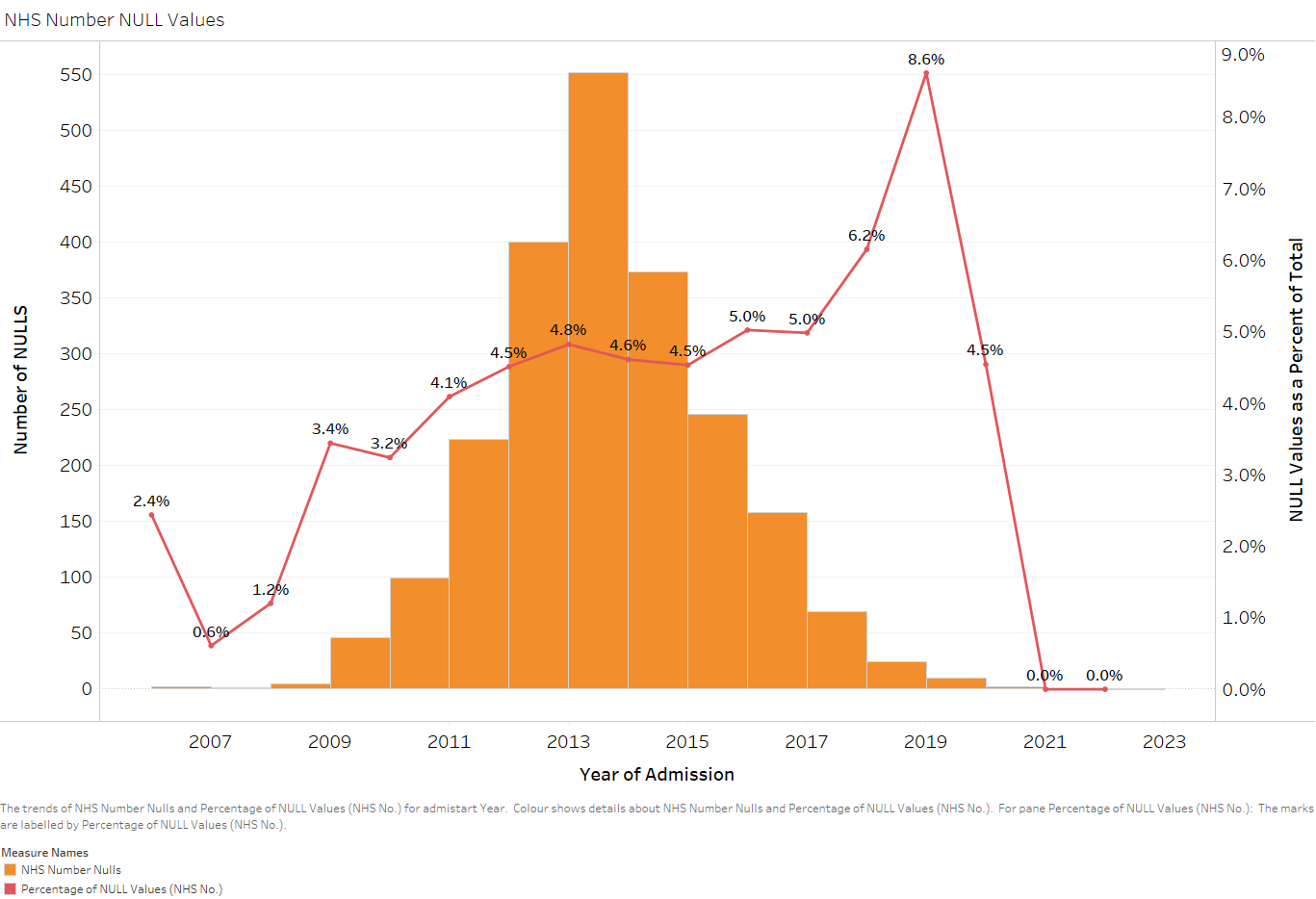
DG Dimension: **Completeness**. There were 6% nulls or invalid values within Ethnos indicating 94% completeness.



Legal Category:

Summary: This shows the number of valid legal categories compared to results which were entered as ‘Not Valid’ or ‘Not Known’.

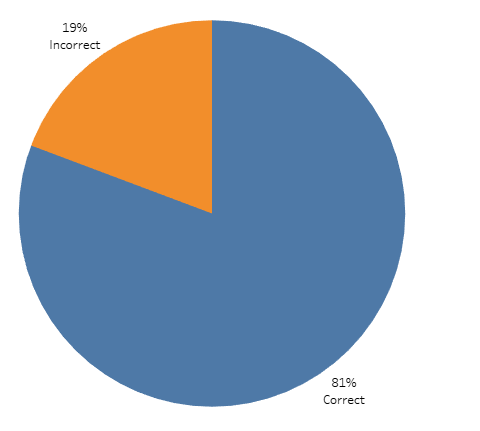
DQ Dimension: **Validity**. The graph shows there is 99.5% validity.



NHS Number

Summary: The graph above shows the number of null values for each year as a percentage of the total

Dimension: Completeness, the number of nulls is an inverse measure of Completeness. E.g. in 2019 the % of nulls was 8.6% so the completeness was 91.4%.



Epidur

Summary**:** The chart above indicates the percentage of values in the field epidur which are correct based on a comparison with the calculation of difference in day between the between the beginning and end of the episode.

Dimension: Accuracy. The % of correct entries is a measure of accuracy. For this field, accuracy is 81%.