

Wikidata by country

The Wealth of Nation's Data

Introduction

This is a quick analysis of country-related statistics of Wikidata items, from November 2014 to April 2015. Observed were citizens (“nationality”) and locations.

Observed countries were Australia (AUS), France (FR), New Zealand (NZ), and the United Kingdom (UK); these were chosen primarily because they can be reasonably well covered with “a circle on a map” (circular geographic area), which allows comparison of country-by-coordinate vs. country-by-administrative unit.

Data was collected daily, using WDQ and Labs databases. Data collection suffered problems for all countries except UK, for December 2014 and January 2015. The cause is unknown, though I suspect a Labs database timeout problem. The data collection is ongoing. Live data can be found [here](#), the raw data used for this analysis is [here](#).

Results

Citizens

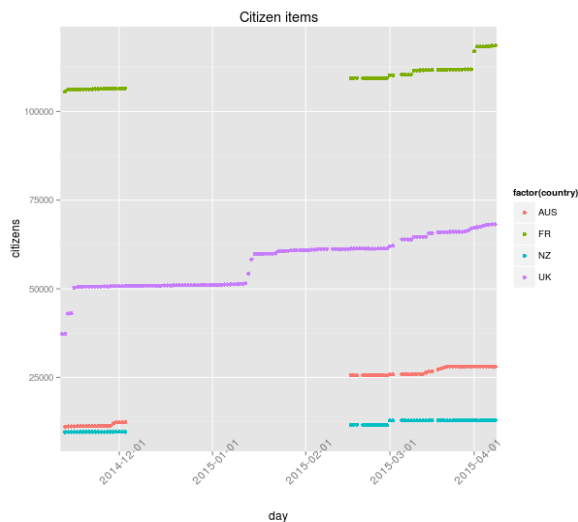


Figure 1.

This plot shows the total number of citizen items per country on Wikidata. Despite a similar population size (~60-65 million), France has about twice as many citizen items.

UK saw a sharp increase in numbers in mid-January, possibly due to the mass-creation of items from the ODNB, or a refresh in the WDQ dataset.

Australia more than doubled its citizen items between December and mid-February. New Zealand shows a slight, gradual increase.

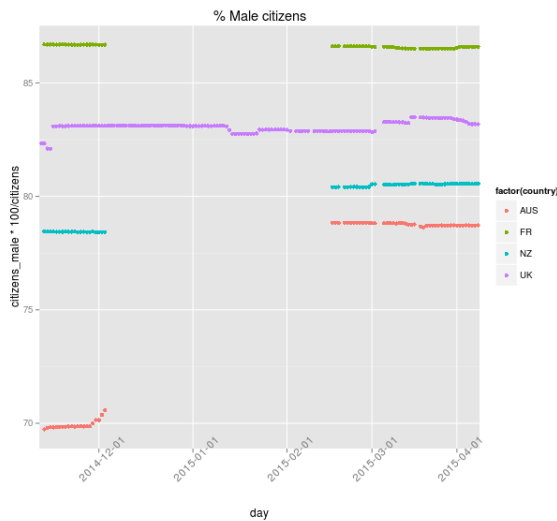


Figure 2.

Both France and the UK show a consistent percentage of male citizens, with France taking the chauvinist lead at 87%, followed by the UK at 83%.

New Zealand shows a small, Australia a significant (70% to 79%) increase in male citizens. These changes may be due to the relatively low number of citizen items.

These numbers are reflected in the percentage of female citizens (plot not shown), and not due to genderless items.

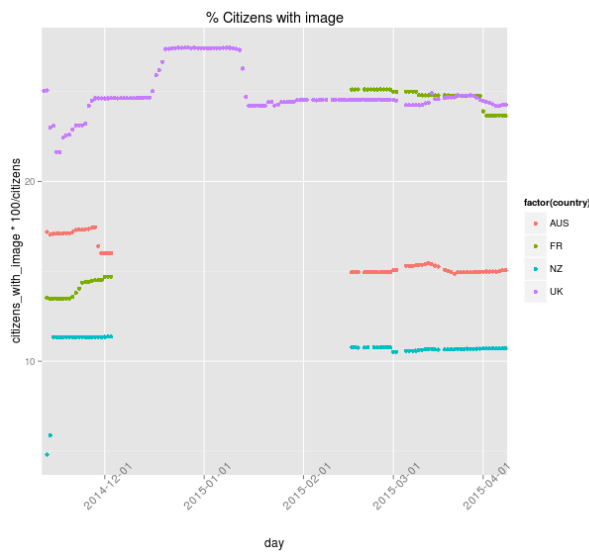


Figure 3.

For both France and the UK, about ¼ of their citizen items have an image.

The UK had a bump in mid-December, pushing towards 30%, which was then offset by the influx of new, image-less items in mid-January (see Fig. 1).

France started below 15%, but caught up to the UK between December and February.

Both Australia and New Zealand stayed at roughly the same level, respectively. A small drop for Australia can be explained by the slight increase in total citizen items (see Fig. 1).

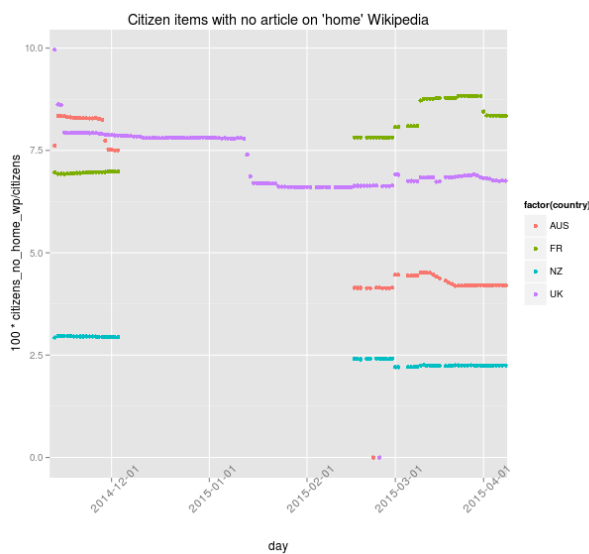


Figure 4.

This plot shows the percentage of citizen items without an entry on their respective “home” Wikipedia (FRWP for France, ENWP otherwise).

The UK shows a small, Australia a steep decline in citizen items without an ENWP article. The UK drop is surprising, as the total number of citizen items increased at the same point in time (see Fig. 1). France shows a small increase in “homeless” citizen items.

Locations

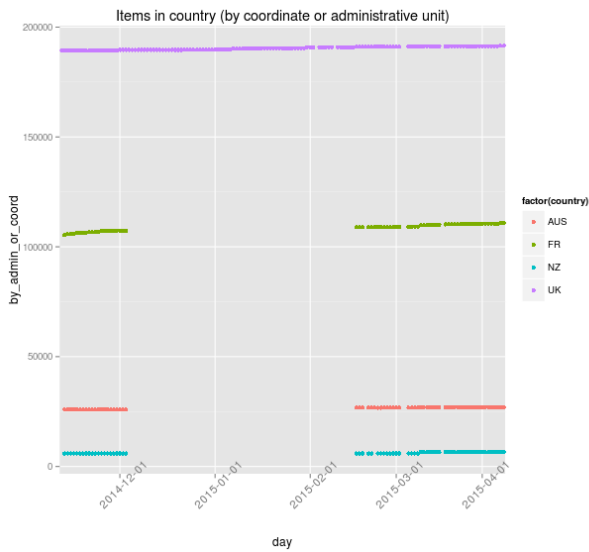


Figure 5.

This plot shows the number of items that are either in an “administrative unit” within the respective item tree of the country, or within the geographic region as defined by one or more circular regions (overseas territories not included).

The number of items stays roughly constant in each country, with a slight upwards trend, which is to be expected, as new items are created and existing items are improved.

Despite having only half the surface (243,600 km² in the UK vs. 551,500 km² in France), the UK has almost twice as many items than France, making it much more densely covered on Wikidata.

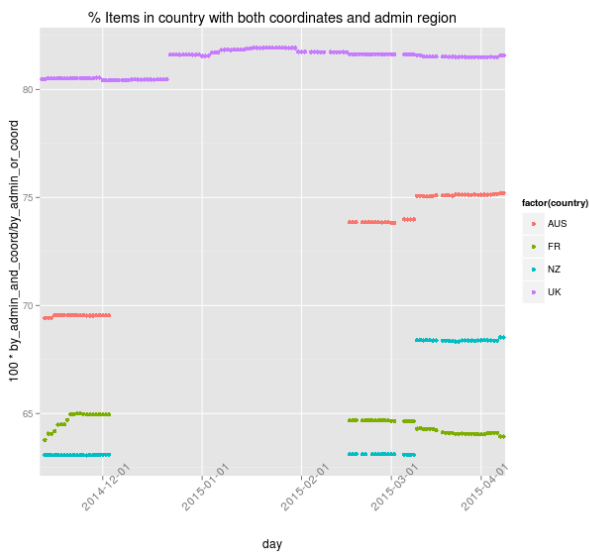


Figure 6.

This plot shows the percentage of items having both coordinates and administrative region, using items with either/or as 100%.

UK items are performing best by far, at consistently >80%. Australia and New Zealand do less well, but both show significant improvement. France performs worst, staying consistently below 65%.

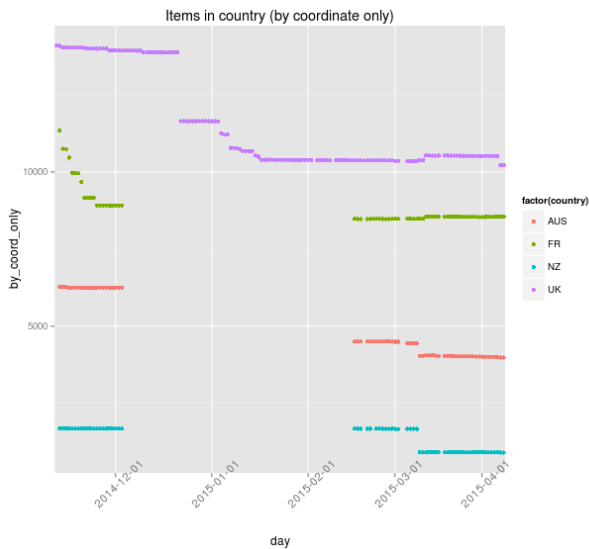


Figure 7.

All countries show a steady decrease in items that have a coordinate, but are not in an administrative region. This corresponds with a steady increase in items that have both (plot not shown). This indicates items being increasingly annotated with an administrative region. This still leaves many tens of thousands of items unlinked to the admin-region tree in these four countries alone. Numbers for the opposite situation (admin-unit but no coordinate) are even higher (plot not shown) and increasing, which is interpreted as a failure of geo-tagging such items with a coordinate.

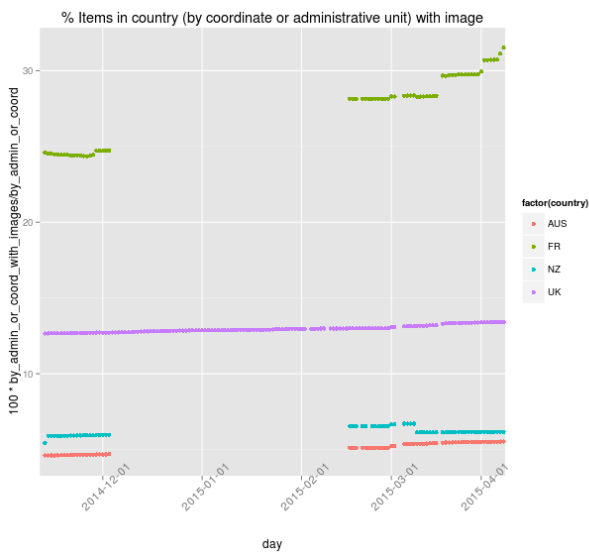


Figure 8.

While the UK has many more Wikidata items, France has a much higher percentage of such items with an image. Furthermore, while Australia, New Zealand, and the UK show only a low, gradual increase in images, the ratio of items with images for France has drastically increased over the last few month, and is now at over 30%, compared to ~13% for the UK. This is despite the dense image coverage of UK images from geograph.co.uk.

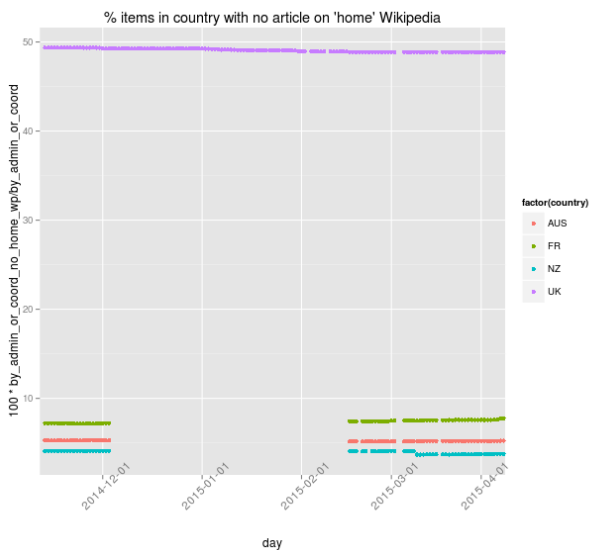


Figure 9.

This plot shows the percentage of items in a country (by either coordinate or administrative region) without a corresponding article in their “home” Wikipedia (FRWP for France, ENWP otherwise).

Surprisingly, while ~95% of items from Australia, France, and New Zealand have a “home” Wikipedia article, almost half (or ca. 100,000) of the UK items do not. This can partially be explained by the creation of Wiki Loves Monuments UK heritage items.

Results

Despite a similar current population size, France has twice as many citizen items than the UK. France has the most male items, Australia the least. Both France and the UK have images for $\sim\frac{1}{4}$ of their citizen items. All countries have articles on their respective home language Wikipedia for $>90\%$ of their citizen items.

Despite having twice the area, France has only $\frac{1}{2}$ the number of location items of the UK, giving the UK four times the item density of France. However, about half of the UK location items have no corresponding article on English Wikipedia, implying they were created in bulk from third-party sources (e.g. National Heritage for WLM). The UK has the highest percentage ($>80\%$) of location items with both administrative region and coordinates; France has the lowest percentage ($<65\%$). In contrast, 30% of French location items have an image, whereas only 15% of UK location items have one; however, adjusting for the UK “third-party items”, this might even out.