CONFERENCE OF EUROPEAN STATISTICIANS

Workshop on Statistical Dissemination and Communication WP.4-5

12-14 June 2019, Gdansk, Poland 07 June 2019

Digging deeper into web analytics. Getting google to work for statistics offices

Mikel Bilbao (EUSTAT – Basque Statistics Office)

mikel@eustat.eus

Abstract

Custom variables in google analytics

Google analytics is the de-facto industry standard when it comes to measuring web products' performance. Out of the box, it gives a ton of useful information to the knowledgeable webmaster, but it can be customize for special needs. And statistics offices do have special needs.

Measuring the performance of the whole website, or a single page or certain subsets is not complicated but, what about the performance of a specific survey and all of its products? Or what about the press releases, all of them? Do you tell apart the success of the statistical products and the navigation pages?

There are plenty of interesting questions that need complicated and time consuming processes if analytics is used as is. This is where customizing/tagging comes into play.

In Eustat we are using custom variables to tag every page of our website. These variables are relevant to statistical production (survey code, product type) or dissemination (format, language) plus a fifth we currently use as an auxiliary variable for different purposes depending on context.

Here an example of the analytics script customization:

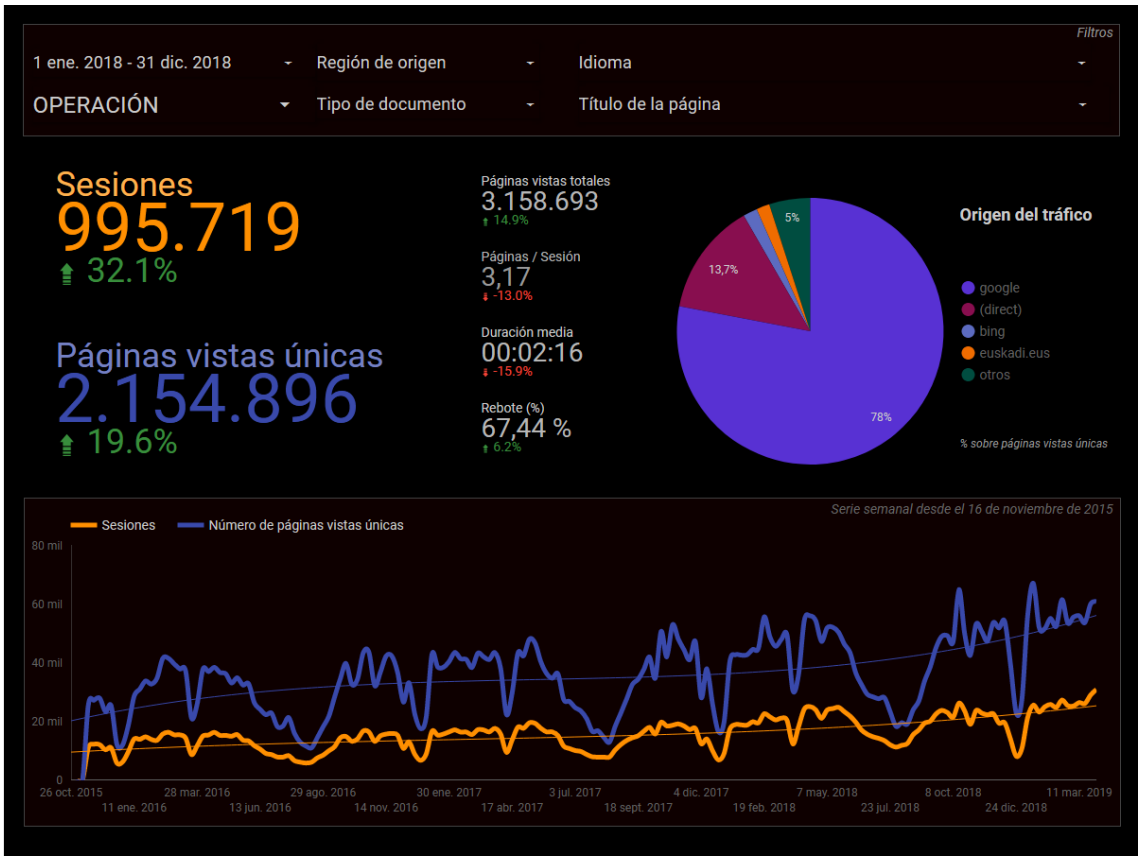
```
<script type='text/javascript'>
     (function (i, s, o, g, r, a, m) {
         i['GoogleAnalyticsObject'] = r; i[r] = i[r] || function () {
             (i[r].q = i[r].q || []).push(arguments)
         }, i[r].l = 1 * new Date(); a = s.createElement(o),
            m = s.getElementsByTagName(o)[0]; a.async = 1; a.src = g; m.parentNode.insertBefore(a, m)
     })(window, document, 'script', '//www.google-analytics.com/analytics.js', 'ga');
     ga('create', 'DUMMY', 'auto');
        ga('set', 'dimension1', 'C');          Language (C for Spanish)
        ga('set', 'dimension2', '56');         Aux. info (Topic, in this case)
        ga('set', 'dimension3', '112311');     Survey code (Offcial code)
        ga('set', 'dimension4', 'NOTAPRENSA'); Type of document (NOTAPRENSA stands for "Press release")
        ga('set', 'dimension5', 'ASP');        Format (Standard web: ASP)
     ga('send', 'pageview');
</script>
```

AzterWEB – web performance reporting dashboard

As a side benefit (not at all small), our customization of analytics allowed us to build a dashboard -a web app-, available to everyone within the Basque Statistics Organization to consult the performance statistics of the entire website or filter according to one's special needs. This way, survey's technicians can check the most relevant pieces of data, with the level of detail they desire, from the whole website to a single document.

This dashboard could even be made public, if the organization so desire, as a means for transparency.

This is how it looks without any filtering applied.



Digging deeper into web analytics. Getting google to work for statistics offices

Mikel Bilbao (EUSTAT – Basque Statistics Office)
mikel@eustat.eus

Introduction

Measuring the success (or failure) of our efforts to disseminate and make statistics popular is key for every statistics office. When it comes to web-based products' performance analytics tools make this task easy enough for any web master.

Google analytics -the de-facto industry standard- provides a comprehensive set of tools that could even be overkill for most organizations. The granularity allowed by analytics, out of the box, is overwhelming. It could even be the base for a full-time job.

Having said that, analytics won't be enough to meet some special needs a statistical office could have. For instance, it would be painfully laborious filtering all the products of a single survey in order to adding them up and measure the performance of the survey on the whole. Not to mention, making it for all the available surveys.

How do our press releases perform? And our data banks? Languages? Data bank vs. pre-formatted tables for the same survey? PDF vs DOC vs HTML?

Questions like these can be answered via painstaking filtering but there are better ways to do it. URL based tagging or utm campaign tracking are two of these ways, but the best we know of is, by a long shot, using "custom variables".

Custom variables in google analytics make it possible to answer all these questions the easy way and, to make things better, they can be used as native variables on analytics mixing them in any report.

Even more, they make it possible to set up live statistics boards (via google data studio) so that all that knowledge can be shared with all the organization (or even make it public if so desired) as we'll see in the second part of this paper.

Chapter 1 – Measuring success

Scenario

In our statistical office we are measuring web traffic with google analytics. Plenty happy with it but, being it a one-size-fits-all service, it lacks variables we, as statistical office, need. Being "Survey" the most important one.
There's more than one way to achieve this. It can be done inserting the information in the URLs and filtering in our analytics tool; it can be done via "utm" campaign URL marking, but the best, most useful and flexible way we know of is by using custom variables.

Custom variables in google analytics

Custom variables –according to google itself- are:
"name-value pair tags that you can insert in your tracking code in order to refine Google Analytics tracking. With custom variables, you can define additional segments to apply to your visitors other than the ones already provided by Analytics."
For our scenario "aggregations" fits better than "segments" and "pages" fits better than "visitors" as we are measuring the performance of our products and not segmenting our visitors.
Custom variables, then, allow us to make aggregations that match our needs, out of the standards google provide.
But, it's important to know that they can be used also at the visitor level, for example, to track paying and non-paying users (in case the website offers premium content); and they can be used at the session level, for example, to segment logged and non-logged visitors.

5 custom variables

Analytics offers five slots you can populate with your custom variables. We used them all, of course.
For our needs, we've set up these five:

- **Language** of the content (Spanish, Basque, English)
- **Auxiliary** for many needs (formerly it was intended for **"topic"**)
- **Survey code** (official six-digit code that identifies our surveys)
- **Product type** (e.g. Press release, table, data bank, definition, graph, report, classification, form…)
- **Format** (e.g. html, xls, pdf, px, doc,…)

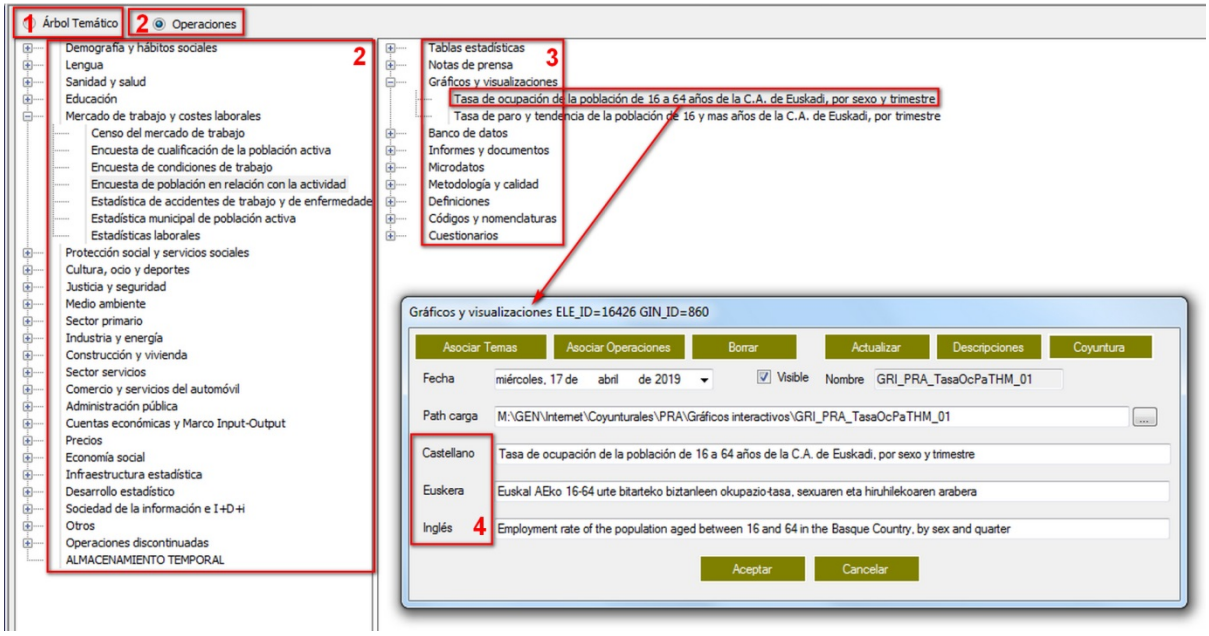
And it looks like this on the source code of one of our pages:

```
<script type='text/javascript'>
     (function (i, s, o, g, r, a, m) {
         i['GoogleAnalyticsObject'] = r; i[r] = i[r] || function () {
             (i[r].q = i[r].q || []).push(arguments)
         }, i[r].l = 1 * new Date(); a = s.createElement(o),
            m = s.getElementsByTagName(o)[0]; a.async = 1; a.src = g; m.parentNode.insertBefore(a, m)
     })(window, document, 'script', '//www.google-analytics.com/analytics.js', 'ga');
     ga('create', 'DUMMY', 'auto');
        ga('set', 'dimension1', 'C');        → Language ("C" for "Spanish")
        ga('set', 'dimension2', '116');      → Topic (116 for "Industry")
        ga('set', 'dimension3', '112312');   → Survey code (112312 for "Industrial production index")
        ga('set', 'dimension4', 'NOTAPRENSA'); → Product type (NOTAPRENSA for "Press release")
        ga('set', 'dimension5', 'ASP');      → Format (ASP for "web standard")
     ga('send', 'pageview');
</script>
```

Looking at this code I already know we are before a press release, in Spanish, html view and both the topic (113) and survey code (112312) correspond to the Industrial production index. Of course, the power of this code relays in the fact that every Spanish page is tagged the same, so they can be aggregate; every press release is tagged the same, so they can be aggregate; every Industrial production index document is tagged the same, so they can be aggregate and so on.

CMS populates it with proper data

Certainly, we do not tag manually every page we disseminate. Our CMS does the heavy lifting:

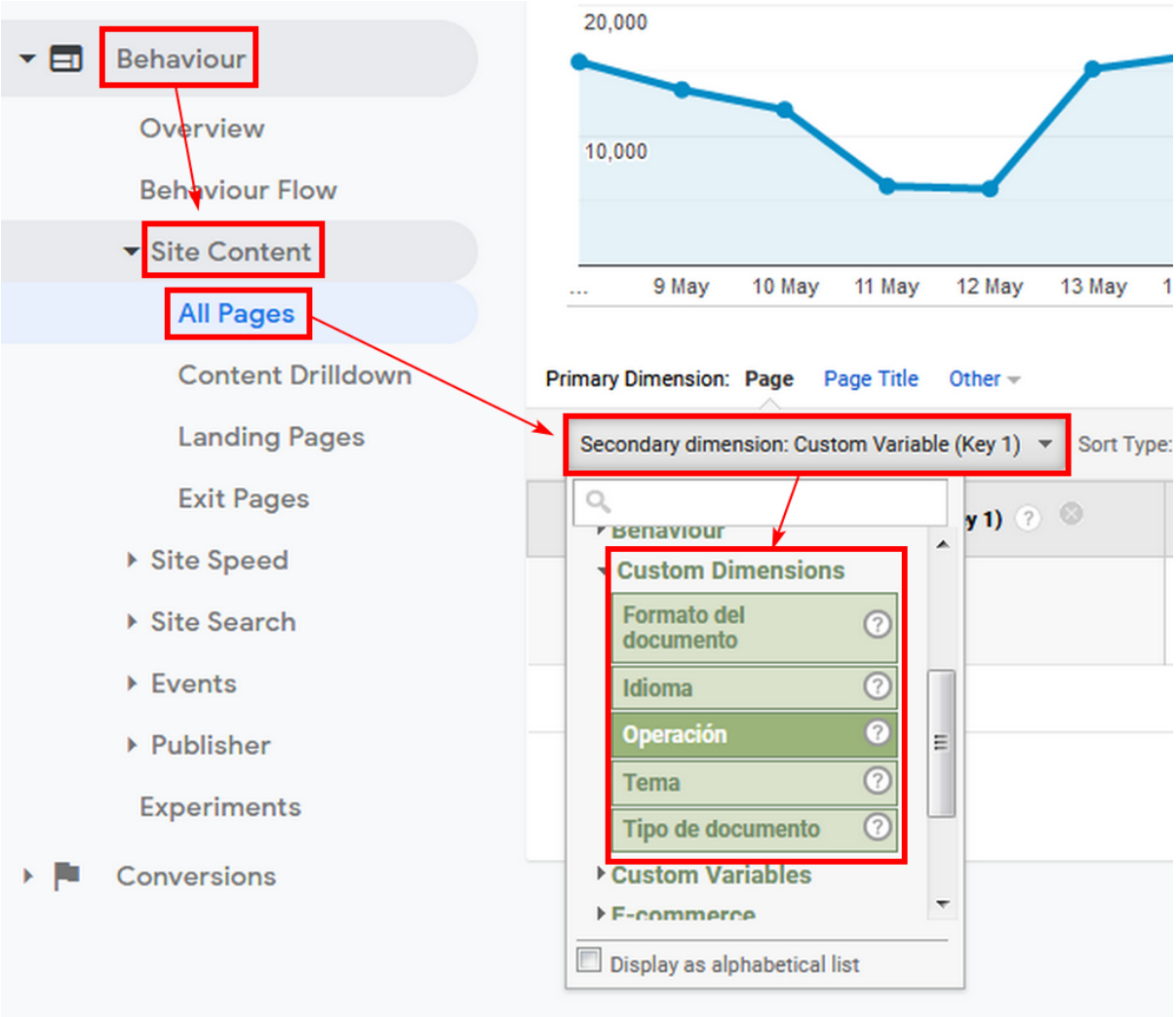


In the process of uploading a product to our website we must assign
1.- Topic
2.- Survey
3.- Type of product
4.- Language

And all this information, along with the final format, is used to populate the custom variables of the analytics script within the webpage...

Harvesting on analytics

Now that we have seeded the land, and the data has grown, it's harvesting time.
On analytics, we can find our own data ready to be used on analyzing or reporting:



And we get this:

	Page Title	Operación	Page Views	Unique Page Views	Avg. Time on Page
		Survey code	313,198 % of Total: 95.69% (327,314)	219,896 % of Total: 96.33% (228,269)	00:01:06 Avg for View: 00:01:07 (-1.22%)
1.	Eustat. PXWeb	0	45 (0.01%)	13 (0.01%)	00:01:11
2.	Euskal AEk hirugarren dirau herrialdeen 2007ko Giza Garapenaren Indizean	010106	1 (0.00%)	1 (0.00%)	00:00:00
3.	Euskal AEk munduko maila altuenetakoa giza garapenean	010106	1 (0.00%)	1 (0.00%)	00:00:00
4.	Giza garapenaren indizea adierazleen arabera herrialdeei jarraiki. 2018	010106	59 (0.02%)	47 (0.02%)	00:03:36
5.	Giza garapenaren indizea dimentsioen, adierazleen eta indizeen arabera. 2010-2017	010106	3 (0.00%)	2 (0.00%)	00:13:54
6.	In 2017 the Basque Country remained amongst the top 15 countries according to the HDI ranking	010106	2 (0.00%)	2 (0.00%)	00:00:15
7.	Índice de Desarrollo Humano de la C.A. de Euskadi por dimensiones, indicadores e índices. 2010-2017	010106	41 (0.01%)	41 (0.02%)	00:00:44
8.	Índice de Desarrollo Humano por indicadores según países. 2018	010106	478 (0.15%)	402 (0.18%)	00:04:33
9.	La C. A. de Euskadi se mantiene en el tercer puesto del Índice de Desarrollo Humano de 2007 por países	010106	7 (0.00%)	7 (0.00%)	00:00:00
10.	La C.A. de Euskadi alcanza uno de los índices de desarrollo humano más altos del mundo	010106	3 (0.00%)	3 (0.00%)	00:00:00

	Page Title	Operación	Page Views	Unique Page Views	Avg. Time on Page
		Survey code	313,198 % of Total: 95.69% (327,314)	219,896 % of Total: 96.33% (228,269)	00:01:06 Avg for View: 00:01:07 (-1.22%)
1.	Eustat. PXWeb	0	45 (0.01%)	13 (0.01%)	00:01:11
2.	Euskal AEk hirugarren dirau herrialdeen 2007ko Giza Garapenaren Indizean	010106	1 (0.00%)	1 (0.00%)	00:00:00
3.	Euskal AEk munduko maila altuenetakoa giza garapenean	010106	1 (0.00%)	1 (0.00%)	00:00:00
4.	Giza garapenaren indizea adierazleen arabera herrialdeei jarraiki. 2018	010106	59 (0.02%)	47 (0.02%)	00:03:36
5.	Giza garapenaren indizea dimentsioen, adierazleen eta indizeen arabera. 2010-2017	010106	3 (0.00%)	2 (0.00%)	00:13:54
6.	In 2017 the Basque Country remained amongst the top 15 countries according to the HDI ranking	010106	2 (0.00%)	2 (0.00%)	00:00:15
7.	Índice de Desarrollo Humano de la C.A. de Euskadi por dimensiones, indicadores e índices. 2010-2017	010106	41 (0.01%)	41 (0.02%)	00:00:44
8.	Índice de Desarrollo Humano por indicadores según países. 2018	010106	478 (0.15%)	402 (0.18%)	00:04:33
9.	La C. A. de Euskadi se mantiene en el tercer puesto del Índice de Desarrollo Humano de 2007 por países	010106	7 (0.00%)	7 (0.00%)	00:00:00
10.	La C.A. de Euskadi alcanza uno de los índices de desarrollo humano más altos del mundo	010106	3 (0.00%)	3 (0.00%)	00:00:00

What's more, we can set up a custom report for, for example, types of documents:

Tipo de documento	Users	Sessions	Unique Page Views	Page Views
Document type	1,178,410 % of Total: 101.01% (1,166,593)	1,557,249 % of Total: 99.70% (1,561,941)	3,337,851 % of Total: 99.59% (3,351,691)	4,907,753 % of Total: 99.64% (4,925,310)
1. DEFINICION	**756,517** (46.61%)	811,653 (52.12%)	902,265 (27.03%)	1,028,138 (20.95%)
2. TABLA	**165,779** (10.21%)	160,077 (10.28%)	361,490 (10.83%)	457,330 (9.32%)
3. LISTADO	**135,289** (8.34%)	63,248 (4.06%)	621,417 (18.62%)	1,062,774 (21.66%)
4. VACIO	**115,136** (7.09%)	86,278 (5.54%)	314,202 (9.41%)	464,226 (9.46%)
5. HOME	**99,346** (6.12%)	168,038 (10.79%)	198,928 (5.96%)	294,821 (6.01%)
6. FICHA_MUNICIPAL	**55,857** (3.44%)	63,191 (4.06%)	102,446 (3.07%)	197,123 (4.02%)
7. BANCO_PX	**55,520** (3.42%)	25,115 (1.61%)	388,738 (11.65%)	661,059 (13.47%)
8. FORMULARIO	**45,287** (2.79%)	55,224 (3.55%)	67,945 (2.04%)	95,049 (1.94%)
9. NOTAPRENSA	**39,251** (2.42%)	34,721 (2.23%)	70,967 (2.13%)	86,259 (1.76%)
10. FICHAMETOD	**22,642** (1.39%)	23,830 (1.53%)	26,284 (0.79%)	29,586 (0.60%)
11. ARBOL	**22,393** (1.38%)	4,297 (0.28%)	98,624 (2.95%)	156,718 (3.19%)
12. INFO	**18,962** (1.17%)	11,877 (0.76%)	39,538 (1.18%)	64,304 (1.31%)
13. INTERACTIVO	**15,982** (0.98%)	7,766 (0.50%)	27,721 (0.83%)	34,608 (0.71%)
14. GRAFICO	**13,897** (0.86%)	7,142 (0.46%)	24,381 (0.73%)	37,459 (0.76%)
15. EXPORTADO	**12,246** (0.75%)	457 (0.03%)	3,670 (0.11%)	44,500 (0.91%)
16. INTERMEDIA	**12,050** (0.74%)	8,919 (0.57%)	25,172 (0.75%)	48,164 (0.98%)
17. TEST	**10,393** (0.64%)	11,294 (0.73%)	14,634 (0.44%)	40,565 (0.83%)
18. WEBESCOLAR	**5,734** (0.35%)	5,658 (0.36%)	10,166 (0.30%)	16,058 (0.33%)
19. UDIDACTICAS	**3,006** (0.19%)	1,169 (0.08%)	8,065 (0.24%)	15,510 (0.32%)
20. CENSOHIST	**2,316** (0.14%)	227 (0.01%)	3,069 (0.09%)	4,578 (0.09%)
21. APARTADO_MUNICIPAL	**1,912** (0.12%)	652 (0.04%)	2,505 (0.08%)	4,458 (0.09%)
22. JUEGOS	**1,780** (0.11%)	1,033 (0.07%)	4,335 (0.13%)	30,077 (0.61%)
23. INFORMES_DOCUMENTOS	**1,607** (0.10%)	706 (0.05%)	4,414 (0.13%)	6,947 (0.14%)
24. EJERCICIOS	**1,553** (0.10%)	1,493 (0.10%)	1,786 (0.05%)	3,840 (0.08%)
25. FICHACALIDAD	**1,402** (0.09%)	331 (0.02%)	2,489 (0.07%)	2,728 (0.06%)

And the same goes for any other variable we are interested on.

Chapter 2 – Internal communication

Scenario

People responsible for surveys want to know how their products perform. Most importantly, the direction board want it too.

Over the time we've complied with these demands with a huge biannual report. It's huge because of the scope but the data it provides is very general, very top level.
For each and every survey there's little data (year-on-year evolution of traffic), unless requested for any special reason in which case we provide a custom detailed report on demand.

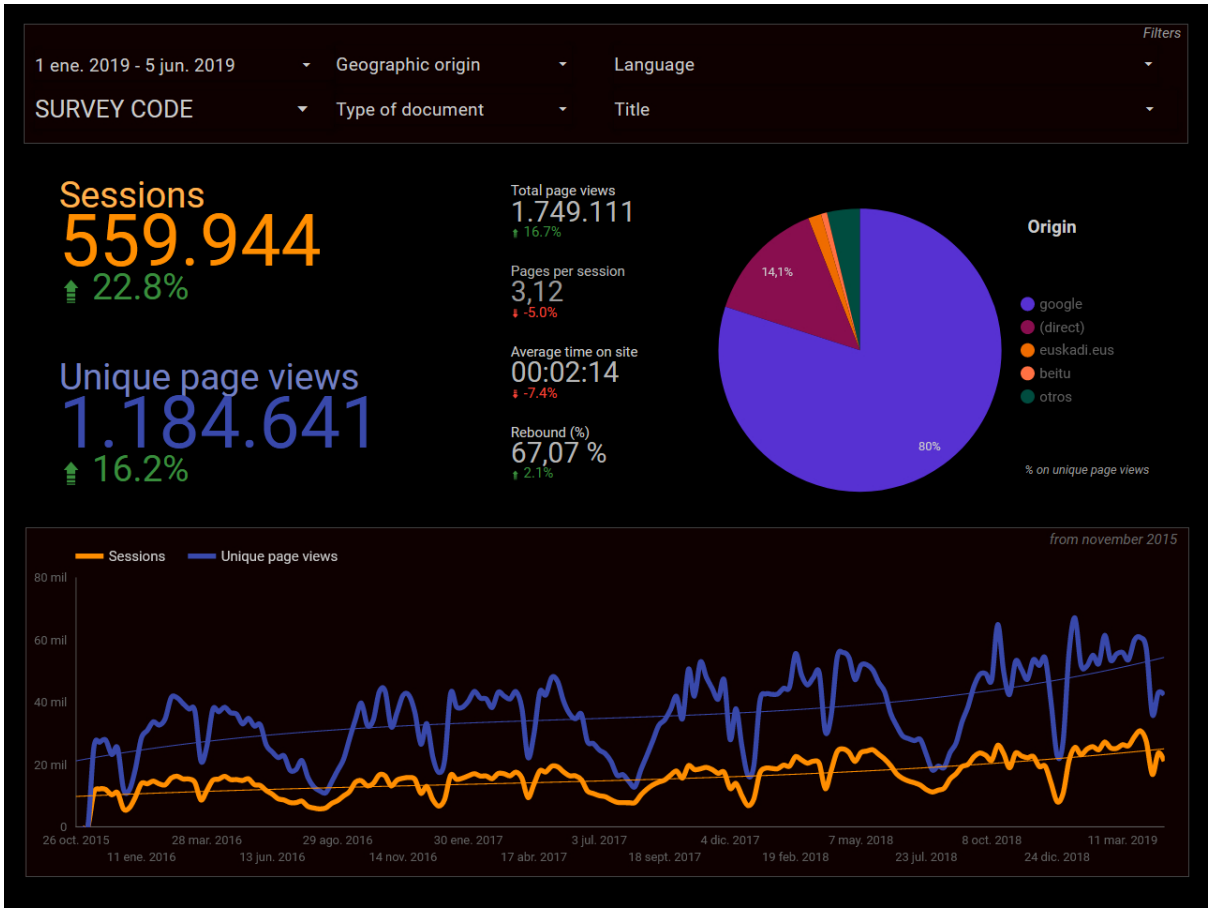
But we found a way to offer even more than they want and with a fraction of the work, thanks to the implementation of custom variables and using Google data studio to set up a statistics dashboard we called AzterWEB ("Azter" being the root of the Basque word "Aztertu" which means "Analyze").

AzterWEB – web performance reporting dashboard

As a side benefit (not at all small), our customization of analytics allowed us to build a dashboard -a web app-, available to everyone within the Basque Statistics Organization to consult the performance of our entire website or filter according to one's special needs. This way, survey's technicians can check the most relevant pieces of data, with the level of detail they desire, from the whole website to a single document.

This dashboard could even be made public, if the organization so desire, as a means for transparency.

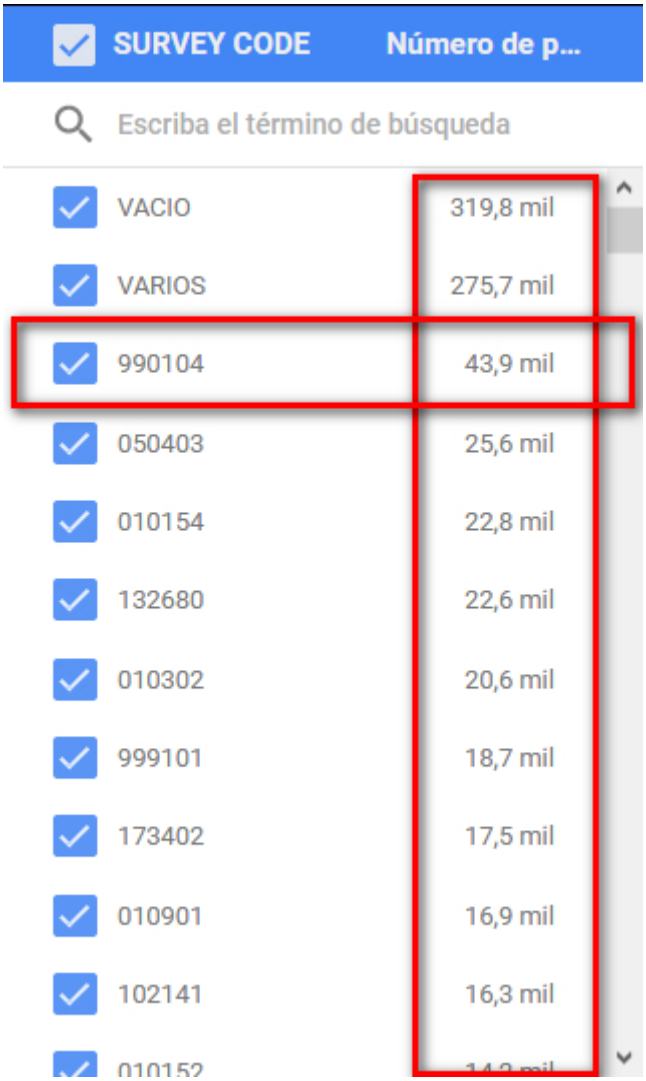
This is how it looks without any filtering applied.



Here we can see the performance of the whole web, this year so far.
But have 6 filters to play around with:

1.- Range of dates
2.- Geographical origin
3.- Language of the content (one of our custom dimensions)
4.- Survey code (custom dimension)
5.- Document type (custom dimension)
6.- Title

Notice the filters themselves are data tables. They provide us with the number of "unique page views" of the categories in a ranking way.



And, it's also interesting how they behave, as the filters filter also the other filters. So, when a survey is selected, the other filters offer only the data of that very survey. That allows any unexperienced analytics user to reach a great level of granularity with ease.