

Calc Beyond Cells

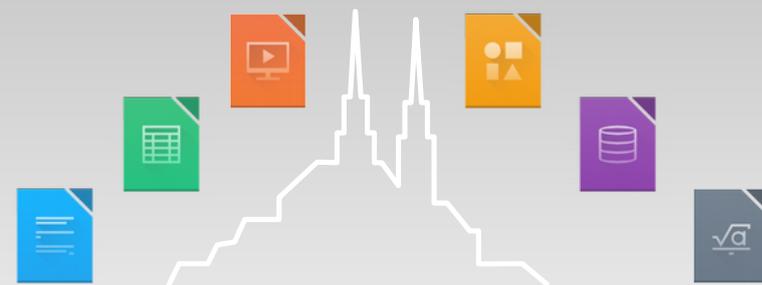
- ▼ Eike Rathke (erAck), Red Hat, Inc.
- ▼ 2016-09-08

About the Speaker



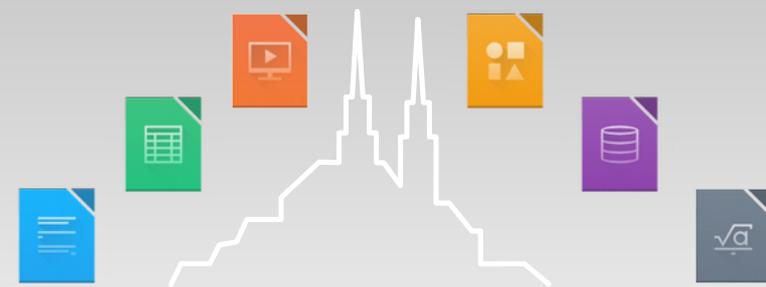
- ▼ Eike Rathke, known on the net as erAck
- ▼ Based in Hamburg, Germany
- ▼ Worked on StarOffice from 1993 to 2000 for Star Division
- ▼ Worked on OpenOffice.org from 2000 to 2011 for Sun Microsystems and one other company
- ▼ Works on LibreOffice since 2011, employed by Red Hat, Inc.
- ▼ Areas of expertise:
 - ▼ Calc core, formula compiler and interpreter
 - ▼ number formatter/scanner
 - ▼ i18n framework, locale data
- ▼ Also mentor and knowledge spreader whenever possible
- ▼ Web site <http://erack.de/>

Agenda



- ▼ There's none..
- ▼ Anyway, what happened since Aarhus 2015
 - ▼ not in terms of release notes, you can read those at
 - <https://wiki.documentfoundation.org/ReleaseNotes/5.1#Calc>
 - <https://wiki.documentfoundation.org/ReleaseNotes/5.2#Calc>
 - ▼ but drawing data from the git repository

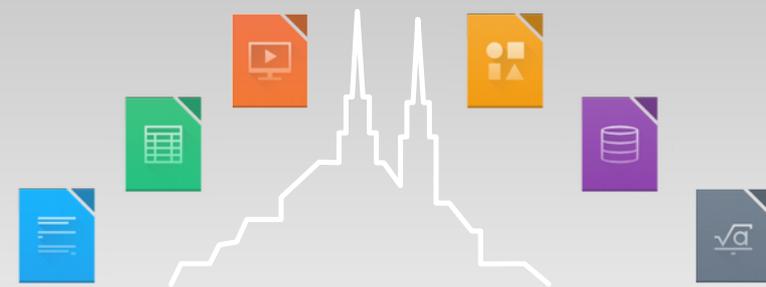
The Long Tail



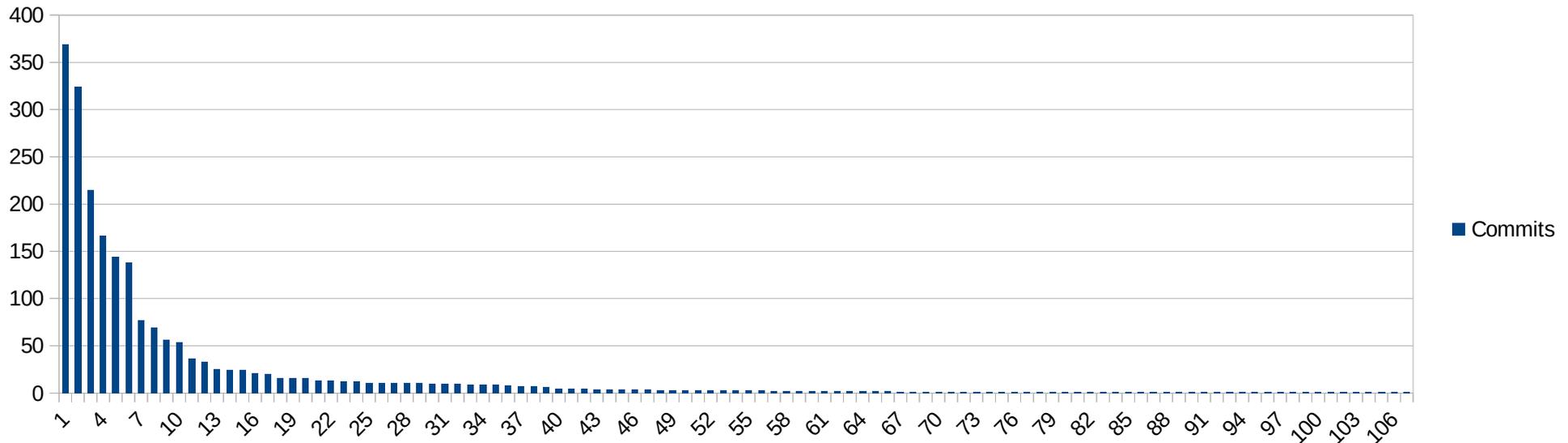
- ▼ All LibreOffice commits from 2015-10-01 to 2016-09-01
 - ▼ 15484 commits by 287 authors
 - ▼ $15484 / 2 = 7742$
 - ▼ the top 5 (1.7% of all) authors contributed 2582, 1858, 1856, 1022 and 633 commits, together 7981 commits, about the half of all commits



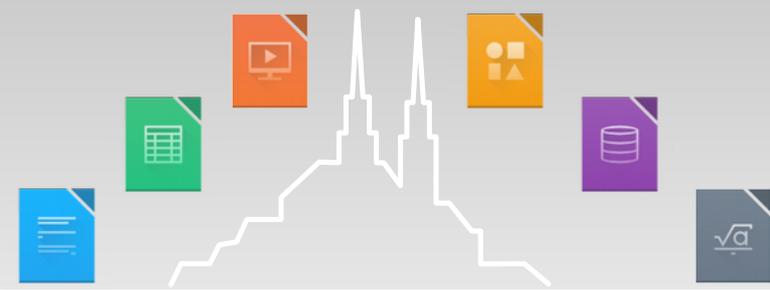
The Long Tail



- ▼ All Calc commits from 2015-10-01 to 2016-09-01
 - ▼ 2158 commits by 107 authors
 - ▼ $2158 / 2 = 1079$
 - ▼ the top 4 (3.7% of all) authors contributed 369, 324, 215 and 167 commits, together 1075 commits, about the half of all commits

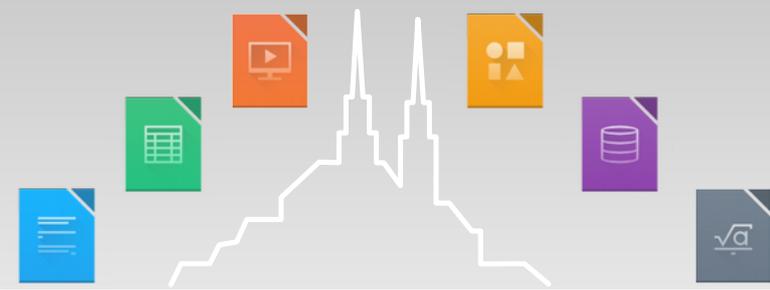


The Fifth Fun Factor



- ▼ The average number of commits per author roughly equals one fifth of the total number of authors
 - ▼ All: 15481 commits by 287 authors
average ~53.95 commits, $287 / 53.95 \approx 5.32$
 - ▼ Calc: 2158 commits by 107 authors
average ~20.17 commits, $107 / 20.17 \approx 5.31$
 - ▼ Writer: 2102 commits by 110 authors
average ~19.11 commits, $110 / 19.11 \approx 5.76$

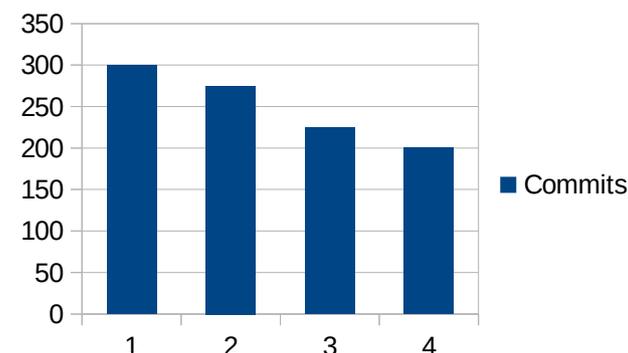
Engagement & Diversity



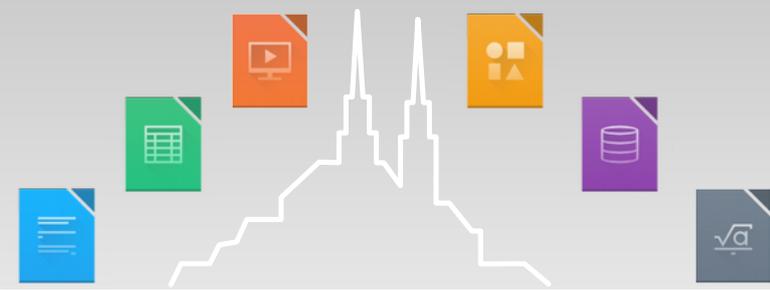
- ▼ The median number of commits: 50% of authors committed more, 50% of authors committed less
- ▼ ratio of average to median, the nearer to 1 the more “conform” or equal the distribution of committers
 - ▼ All: 15484 commits by 287 authors, average ~ 53.95
median 4 commits, ratio $53.95 / 4 \approx 13.49$
 - ▼ Calc: 2158 commits by 107 authors, average ~ 20.17
median 3 commits, ratio $20.17 / 3 \approx 6.72$
 - ▼ Writer: 2102 commits by 110 authors, average ~ 19.11
median 2 commits, ratio $19.11 / 2 \approx 9.55$



- ▼ A hypothetical project with 4 authors and 300, 275, 225 and 200 commits each
 - ▼ the top 2 (50% of all) authors would contribute half of all commits
 - ▼ 1000 commits by 4 authors would give average 250 commits, $4 / 250 = 0.016$
median 250 commits
ratio average to median: $250 / 250 = 1$
- ▼ very low authors per average number
no diversity
⇒ no fun (scratch my itch) factor

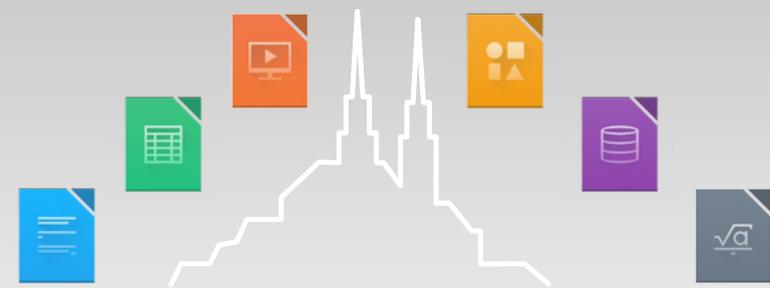


Commit Types in Calc



- ▼ Of 2157 commits
 - ▼ 444 code cleanups and fixups (noelg, sberg)
 - ▼ 42 coverity (coverity#,cid#) (mostly caolan)
 - ▼ 570 bug related (tdf#,rhbz#), including enhancement implementations (77 authors)
 - ▼ 104 commits in sc/qa, bug fix related unit tests
 - ▼ 1087 without bug number (64 authors)
anything from small change to large scale feature
 - ▼ 419 unit test commits in sc/qa (30 authors), of which 219 are new spreadsheet functions tests (raal)

Commits with tdf#



- ▼ 566 commits with tdf#[0-9]+
 - ▼ ranging from [tdf#30456](#) and [tdf#32834](#) to [tdf#101672](#)
 - ▼ tdf#30456 is “enhance Merge Cells dialog to empty cells”
 - ▼ tdf#32834 is “improve precision of MDETERM”
 - ▼ 327 unique bug numbers
 - ▼ top 5:
 - ▼ [tdf#72196](#) with 19 commits, “Implement wildcards” (erack)
 - ▼ [tdf#97831](#) with 15 commits, “New Excel 2016 functions” (winfried)
 - ▼ [tdf#39468](#) with 10 commits, “Translate German comments” (albert)
 - ▼ 4 bug numbers with each 7 commits
 - ▼ 5 bug numbers with each 6 commits

Lines Changed



▼ sc/

- ▼ 58358 lines removed
- ▼ 716840 lines added
- ▼ sc/source/core/
 - ▼ 9889 lines removed
 - ▼ 16013 lines added

▼ sc/source/filter/

- ▼ 10809 lines removed
- ▼ 8676 lines added

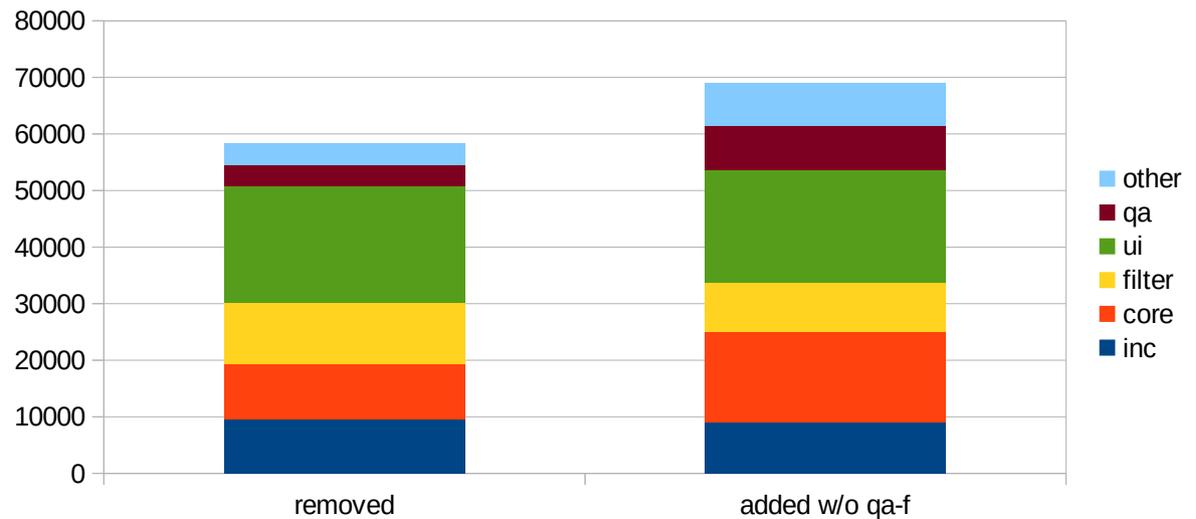
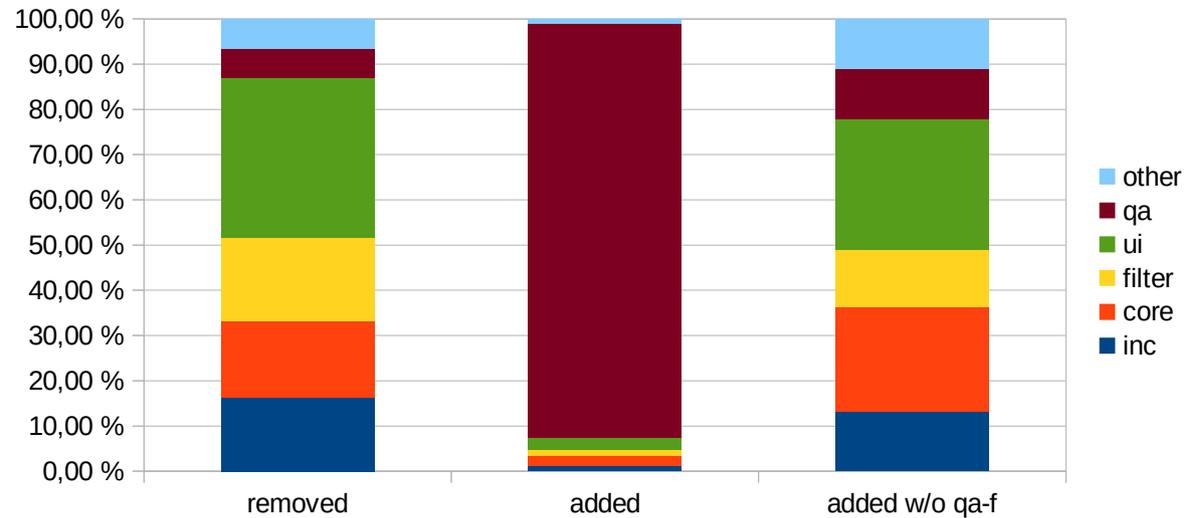
▼ sc/source/ui/

- ▼ 20605 lines removed
- ▼ 19938 lines added

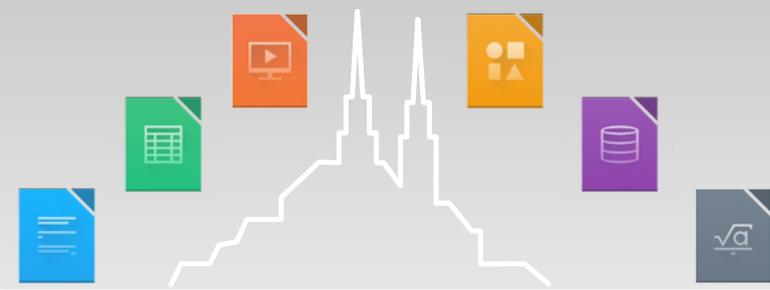
▼ sc/qa/

- ▼ 3700 lines removed
- ▼ 655620 lines added

- ▼ 647814 under sc/qa/unit/data/functions/ (function test documents .fods)



Some Commands Used



▼ Authors sorted by number of commits:

- ▼ `git shortlog --since=2015-10-01T00:00:00 -n sc | grep '^[^].\+:'` | `sed -e 's/^(.\+) (\([0-9]\+\)):^1;\2/'`
 - ▼ import as semicolon separated .csv

▼ Summaries with author;subject:

- ▼ `git log --since=2015-10-01T00:00:00 --pretty='format:%aN;%s' sc`
 - ▼ import as semicolon separated .csv

▼ Authors of a topic:

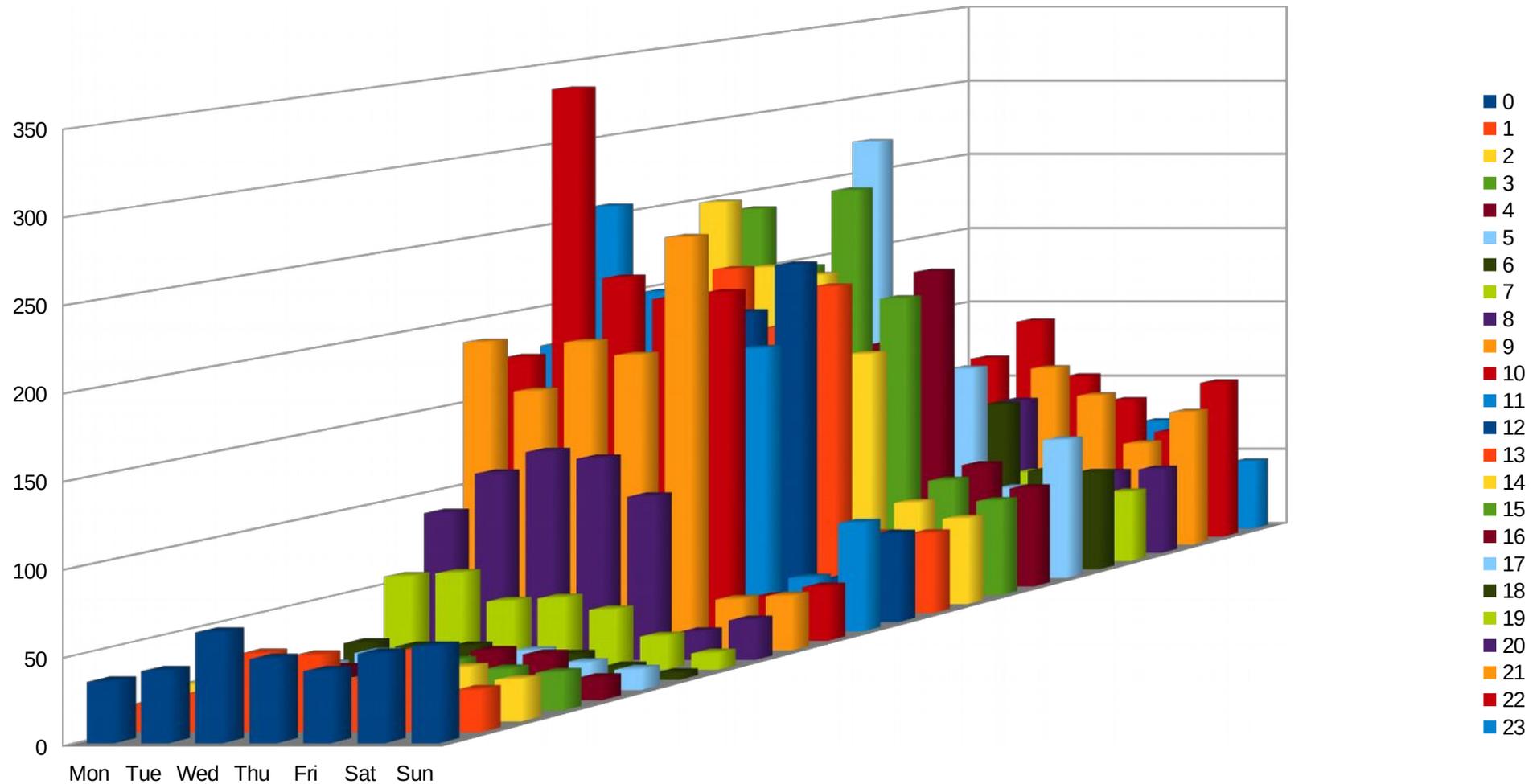
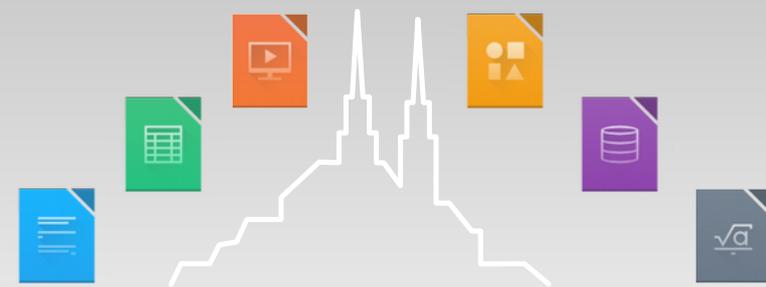
- ▼ `git log --since=2015-10-01T00:00:00 --pretty='format:%aN;%s' sc | egrep '(coverity|cid)#'` | `sed -e 's/^\([^;]*\);.*^1/'` | `sort -u`

▼ Sort commits by tdf# bug numbers:

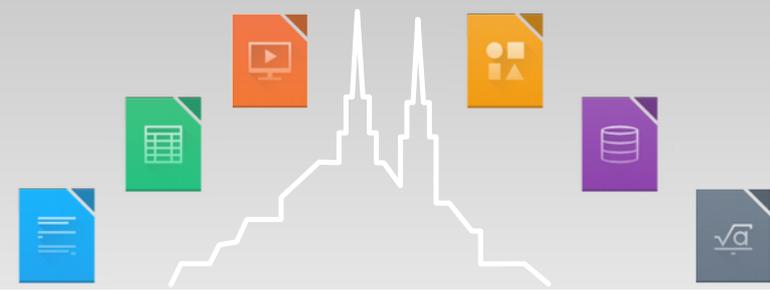
- ▼ `git log --since=2015-10-01T00:00:00 --pretty='format:%aN;%s' sc | grep 'tdf#'` | `sed -e 's/^(.*tdf#)\([0-9]\+\)\(.*\)^2;\1\2\3/'` | `sort -n`
 - ▼ import as .csv and create a pivot table of bug numbers in rows and count

Note: all without .mailmap file, there are ~3 author aliases in Calc and a few more in All.

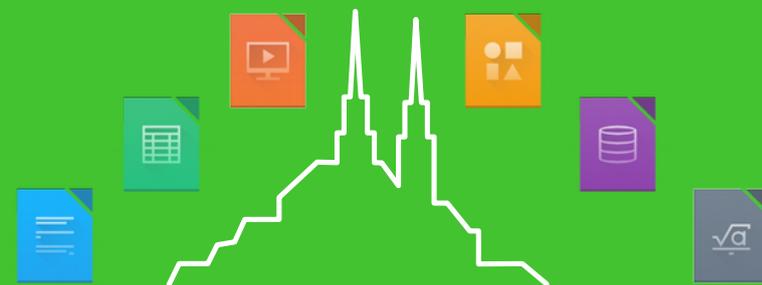
Commits per Hour of Week



Questions?



I might be able to answer..



Thank you ...

- ▾ ... for using LibreOffice!
- ▾ ... for supporting LibreOffice!
- ▾ ... for hacking LibreOffice!



All text and image content in this document is licensed under the [Creative Commons Attribution-Share Alike 3.0 License](https://creativecommons.org/licenses/by-sa/3.0/) (unless otherwise specified). "LibreOffice" and "The Document Foundation" are registered trademarks. Their respective logos and icons are subject to international copyright laws. The use of these therefore is subject to the [trademark policy](#).

