

Program Information:

Language:  
Python

Tools/Libraries:

OS:

This library is used to create subprocesses and pipes. They were used to make git calls with terminal command line and retrieve the output.

Git-Python:

This was also used to make certain git log calls and retrieve the output

JSON:

Json files were created to store the number of files committed together by commit hash. If a Json file was to not be used than the calls would take a while to complete everytime this program is tested.

Itertools:

This library was helpful in ceating combinations of the files committed together. A map-reduce methodology was then used count the number of times a certain file combination was committed to their respective repos.

Numpy, Matplotlib:

This libraries were used to graph the data and display the results using graphs.

---

# all of the output above was produced by the python script provided, assuming all the  
# sub-repos are pulled correctly. They must be in the same directory as this script.  
# with the names commons-math and pdfbox respectively.

Question 1:

Repo: commons-math

a)

Total Number of Commits: 6348

b)

Total Number of Files: 1488

Average Number of Commits Per File =  $6348 / 1488 = 4.266129$

c)

Number of contributors: 55

d)

Average Number of Commits Per Contributor =  $6348 / 55 = 115.418182$

e)

Contributor Name	Contributor Email	Percent of Commits
Rahul Akolkar	rahul@apache.org	0.095%
Ole	ole.ersoy@gmail.com	0.047%
Hang Park	hangpark@kaist.ac.kr	0.016%
Qualtagh	a@a.a	0.032%
Tim O'Brien	tobrien@apache.org	1.134%
Sebastian Bazley	sebb@apache.org	6.648%
Stephen Colebourne	scolebourne@apache.org	0.016%
Brent Worden	brentworden@apache.org	1.197%

Daniil Finkel	dfinkel@coldlight.com	0.016%
Joerg Pietschmann	pietsch@apache.org	0.158%
Eric Barnhill	ericbarnhill@protonmail.c	0.063%
Niall Pemberton	niallp@apache.org	0.536%
Daniel Shahaf	danielsh@apache.org	0.032%
Christoph Dibak	christoph.dibak@ipvs.uni-	0.016%
Dion Gillard	dion@apache.org	0.032%
Thomas Neidhart	tn@tn-desktop.(none)	0.016%
Milos Gligoric	milos.gligoric@gmail.com	0.016%
Evan Ward	evan.ward@nrl.navy.mil	0.047%
Robert Burrell Donkin	rdonkin@apache.org	0.142%
Brett Porter	brett@apache.org	0.016%
Olivier Lamy	olamy@apache.org	0.079%
Emmanuel Bourg	ebourg@apache.org	0.032%
Phil Steitz	phil.steitz@gmail.com	0.945%
Benedikt Ritter	beneritter@gmail.com	0.032%
Hank Grabowski	hank@applieddefense.com	0.079%
Luc Maisonobe	luc@apache.org	29.474%
Rob Tompkins	christopher.tompkins@capi	0.032%
Eric Prescott-Gagnon	eric.prescottgagnon@jda.c	0.016%
Thomas Neidhart	tn@apache.org	6.270%
Phil Steitz	psteitz@apache.org	17.076%
Dirk Verbeeck	dirkv@apache.org	0.252%
Ray DeCampo	ray@decampo.org	0.646%
Artem Barger	artem@bargr.net	0.016%
Dimitri Pourbaix	dimpbx@apache.org	0.142%
Dave Brosius	dbrosius@mebigfatguy.com	0.063%
Luc Maisonobe	luc@orekit.org	0.032%
Thomas Neidhart	thomas.neidhart@gmail.com	1.890%
Henri Yandell	bayard@apache.org	0.504%
William Barker	billbarker@apache.org	0.929%
Mikkel Meyer Andersen	mikl@apache.org	0.441%
No Author	unknown@apache.org	0.016%
Evan Ward	evanward@apache.org	0.032%
Rob Tompkins	chtompki@gmail.com	0.126%
Gilles	erans@apache.org	20.274%
Greg Sterijevski	gregs@apache.org	0.173%
Gary D. Gregory	ggregory@apache.org	0.142%
David Brosius	dbrosius@apache.org	0.047%
darkma773r	matt.juntunen@hotmail.com	0.142%
Sebastien Brisard	celestine@apache.org	6.033%
Gilles	gilles@harfang.homelinux.	0.473%
Matthew Jason Benson	mbenson@apache.org	0.016%
Otmar Ertl	otmar.ertl@gmail.com	0.299%
Bruno P. Kinoshita	brunodepaulak@yahoo.com.b	0.126%
Mark R. Diggory	mdiggory@apache.org	2.867%
Kexin Xie	kexin.xie@salesforce.com	0.016%

f)

Answer Below in Common Repo Answers

g)

Contributors Inactive For More Than 6 Months:

rahul@apache.org  
 ole.ersoy@gmail.com  
 hangpark@kaist.ac.kr  
 a@a.a  
 tobrien@apache.org  
 sebb@apache.org  
 scolebourne@apache.org  
 brentworden@apache.org  
 dfinkel@coldlight.com  
 pietsch@apache.org  
 ericbarnhill@protonmail.ch  
 niallp@apache.org  
 danielsh@apache.org  
 christoph.dibak@ipvs.uni-stuttgart.de  
 dion@apache.org  
 tn@tn-desktop.(none)

milos.gligoric@gmail.com  
 evan.ward@nrl.navy.mil  
 rdonkin@apache.org  
 brett@apache.org  
 olamy@apache.org  
 ebourg@apache.org  
 phil.steitz@gmail.com  
 beneritter@gmail.com  
 hank@applieddefense.com  
 luc@apache.org  
 christopher.tompkins@capitalone.com  
 eric.prescottgagnon@jda.com  
 tn@apache.org  
 psteitz@apache.org  
 dirkv@apache.org  
 artem@bargr.net  
 dimpbx@apache.org  
 dbrosius@mebigfatguy.com  
 luc@orekit.org  
 thomas.neidhart@gmail.com  
 bayard@apache.org  
 billbarker@apache.org  
 mikl@apache.org  
 unknown@apache.org  
 evanward@apache.org  
 chtompki@gmail.com  
 gregs@apache.org  
 ggregory@apache.org  
 dbrosius@apache.org  
 celestin@apache.org  
 gilles@harfang.homelinux.org  
 mbenson@apache.org  
 otmar.ertl@gmail.com  
 brunodepaulak@yahoo.com.br  
 mdiggory@apache.org  
 kexin.xie@salesforce.com

h)

Number Of Files Added: 16  
 Number Of Files Deleted: 5243  
 Number Of Files Modified: 5871

i)

Answer Below in Common Repo Answers

Repo: pdfbox

a)

Total Number of Commits: 6723

b)

Total Number of Files: 1706  
 Average Number of Commits Per File =  $6723 / 1706 = 3.940797$

c)

Number of contributors: 24

d)

Average Number of Commits Per Contributor =  $6723 / 24 = 280.125000$

e)

Contributor Name	Contributor Email	Percent of Commits
Andreas Lehmkuhler	lehmi@apache.org	20.958%
Adam Nichols	adam@apache.org	0.684%
Guillaume Bailleul	gbailleul@apache.org	1.175%
tilman	tilman@unknown	0.312%

Eric Leleu	leleueri@apache.org	1.354%
msahyoun	msahyoun@unknown	0.045%
John Hewson	jahewson@apache.org	11.200%
Philipp Koch	pkoch@apache.org	0.030%
Johannes Koch	koch@apache.org	0.178%
Ben Litchfield	blitchfield@apache.org	0.015%
Thomas Chojecki	tchojecki@apache.org	0.208%
Tim Allison	tallison@apache.org	0.045%
Brian Carrier	carrier@apache.org	0.506%
Maruan Sahyoun	msahyoun@apache.org	6.708%
Villu Ruusmann	vfed@apache.org	0.030%
asf-sync-process	asf-sync-process@unknown	0.119%
jahewson	jahewson@unknown	0.015%
Daniel Wilson	danielwilson@apache.org	0.729%
Jeremias Maerki	jeremias@apache.org	0.535%
Jukka Zitting	jukka@apache.org	3.510%
Timo Boehme	tboehme@apache.org	0.699%
John Hewson	jahewson@apache.org = ja	0.015%
Tilman Hausherr	tilman@apache.org = tilm	0.045%
Tilman Hausherr	tilman@apache.org	50.885%

f)

Answer Below in Common Repo Answers

g)

Contributors Inactive For More Than 6 Months:

adam@apache.org  
 gbailleul@apache.org  
 tilman@unknown  
 leleueri@apache.org  
 msahyoun@unknown  
 jahewson@apache.org  
 pkoch@apache.org  
 koch@apache.org  
 blitchfield@apache.org  
 tchojecki@apache.org  
 tallison@apache.org  
 carrier@apache.org  
 vfed@apache.org  
 asf-sync-process@unknown  
 jahewson@unknown  
 danielwilson@apache.org  
 jeremias@apache.org  
 jukka@apache.org  
 tboehme@apache.org  
 jahewson@apache.org = jahewson = John Hewson jahewson@apache.org@apache.org  
 tilman@apache.org = tilman = Tilman Hausherr tilman@apache.org@apache.org

h)

Number Of Files Added: 73  
 Number Of Files Deleted: 4042  
 Number Of Files Modified: 6462

i)

Answer Below in Common Repo Answers

Common Questions for commons-math and pdfbox :

f. Authors That Contributed to Both Projects:  
 None Found

i : summarize what was observed. Similarities, differences:  
 Observing the results for both repositories showed that just because both repositories belong to apache, it doesn't mean that developers in one project will necessarily contribute to another project. I also observed that only a few developers are typically responsible for the majority of commits in the project. Most contributors are also typically inactive after a certain period of time since

the project was released. After some time only a few contributors continue to push commits, therefore only a few maintain the software.

## Question 2:

commons-math:

a) 4 sets of 2 files that changed together at least 3 times:

Set: 1

```
-src/main/java/org/apache/commons/math3/geometry/euclidean/twod/PolygonsSet.java  
  
-src/test/java/org/apache/commons/math3/geometry/euclidean/twod/PolygonsSetTest.java
```

Set: 2

```
-src/main/java/org/apache/commons/math/analysis/interpolation/MicrosphereInterpolatingFunction.java  
  
-src/main/java/org/apache/commons/math/analysis/interpolation/MicrosphereInterpolator.java
```

Set: 3

```
-src/java/org/apache/commons/math/linear/QRDecomposition.java  
-src/java/org/apache/commons/math/linear/QRDecompositionImpl.java
```

Set: 4

```
-src/main/java/org/apache/commons/math3/fitting/leastquares/GaussNewtonOptimizer.java  
  
-src/test/java/org/apache/commons/math3/fitting/leastquares/AbstractLeastSquaresOptimizerAbstractTest.java
```

b) 3 sets of 3 files that changed together at least 3 times:

Set: 1

```
-src/main/java/org/apache/commons/math/analysis/integration/LegendreGaussIntegrator.java  
  
-src/main/java/org/apache/commons/math/analysis/integration/RombergIntegrator.java  
  
-src/main/java/org/apache/commons/math/analysis/integration/TrapezoidIntegrator.java
```

Set: 2

```
-src/changes/changes.xml  
-src/main/java/org/apache/commons/math3/util/Precision.java  
-src/test/java/org/apache/commons/math3/util/PrecisionTest.java
```

Set: 3

```
-src/java/org/apache/commons/math/StoreUnivariateImpl.java  
-src/java/org/apache/commons/math/Univariate.java  
-src/java/org/apache/commons/math/UnivariateImpl.java
```

c) 2 sets of 4 files that changed together at least 3 times:

Set: 1

```
-src/main/java/org/apache/commons/math/analysis/integration/LegendreGaussIntegrator.java  
  
-src/main/java/org/apache/commons/math/analysis/integration/RombergIntegrator.java  
  
-src/main/java/org/apache/commons/math/analysis/integration/SimpsonIntegrator.java
```

-src/main/java/org/apache/commons/math/analysis/integration/UnivariateRealIntegrator.java

Set: 2

-src/main/java/org/apache/commons/math/analysis/integration/LegendreGaussIntegrator.java

-src/main/java/org/apache/commons/math/analysis/integration/RombergIntegrator.java

-src/main/java/org/apache/commons/math/analysis/integration/SimpsonIntegrator.java

-src/main/java/org/apache/commons/math/analysis/integration/UnivariateRealIntegratorImpl.java

d) 1 sets of 5 files that changed together at least 3 times:

Set: 1

-src/main/java/org/apache/commons/math/analysis/integration/LegendreGaussIntegrator.java

-src/main/java/org/apache/commons/math/analysis/integration/RombergIntegrator.java

-src/main/java/org/apache/commons/math/analysis/integration/TrapezoidIntegrator.java

-src/main/java/org/apache/commons/math/analysis/integration/UnivariateRealIntegrator.java

-src/main/java/org/apache/commons/math/analysis/integration/UnivariateRealIntegratorImpl.java

pdfbox:

a) 4 sets of 2 files that changed together at least 3 times:

Set: 1

-src/main/java/org/apache/commons/math3/geometry/euclidean/twod/PolygonsSet.java

-src/test/java/org/apache/commons/math3/geometry/euclidean/twod/PolygonsSetTest.java

Set: 2

-src/main/java/org/apache/commons/math/analysis/interpolation/MicrosphereInterpolatingFunction.java

-src/main/java/org/apache/commons/math/analysis/interpolation/MicrosphereInterpolator.java

Set: 3

-src/java/org/apache/commons/math/linear/QRDecomposition.java

-src/java/org/apache/commons/math/linear/QRDecompositionImpl.java

Set: 4

-src/main/java/org/apache/commons/math3/fitting/leastquares/GaussNewtonOptimizer.java

-src/test/java/org/apache/commons/math3/fitting/leastquares/AbstractLeastSquaresOptimizerAbstractTest.java

b) 3 sets of 3 files that changed together at least 3 times:

Set: 1

-src/main/java/org/apache/commons/math/analysis/integration/LegendreGaussIntegrator

or.java

-src/main/java/org/apache/commons/math/analysis/integration/RombergIntegrator.java

-src/main/java/org/apache/commons/math/analysis/integration/TrapezoidIntegrator.java

Set: 2

-src/changes/changes.xml

-src/main/java/org/apache/commons/math3/util/Precision.java

-src/test/java/org/apache/commons/math3/util/PrecisionTest.java

Set: 3

-src/java/org/apache/commons/math/StoreUnivariateImpl.java

-src/java/org/apache/commons/math/Univariate.java

-src/java/org/apache/commons/math/UnivariateImpl.java

c) 2 sets of 4 files that changed together at least 3 times:

Set: 1

-src/main/java/org/apache/commons/math/analysis/integration/LegendreGaussIntegrator.java

-src/main/java/org/apache/commons/math/analysis/integration/RombergIntegrator.java

-src/main/java/org/apache/commons/math/analysis/integration/SimpsonIntegrator.java

-src/main/java/org/apache/commons/math/analysis/integration/UnivariateRealIntegrator.java

Set: 2

-src/main/java/org/apache/commons/math/analysis/integration/LegendreGaussIntegrator.java

-src/main/java/org/apache/commons/math/analysis/integration/RombergIntegrator.java

-src/main/java/org/apache/commons/math/analysis/integration/SimpsonIntegrator.java

-src/main/java/org/apache/commons/math/analysis/integration/UnivariateRealIntegratorImpl.java

d) 1 sets of 5 files that changed together at least 3 times:

Set: 1

-src/main/java/org/apache/commons/math/analysis/integration/LegendreGaussIntegrator.java

-src/main/java/org/apache/commons/math/analysis/integration/RombergIntegrator.java

-src/main/java/org/apache/commons/math/analysis/integration/TrapezoidIntegrator.java

-src/main/java/org/apache/commons/math/analysis/integration/UnivariateRealIntegrator.java

-src/main/java/org/apache/commons/math/analysis/integration/UnivariateRealIntegratorImpl.java