git log/history

References

- Visualizing Code: Polyglot Notebooks Repository (YouTube)
- gitstractor (GitHub)
- Neo4j Python Driver

```
Command line execution (CLI mode): Yes
```

Git History - Directory Commit Statistics

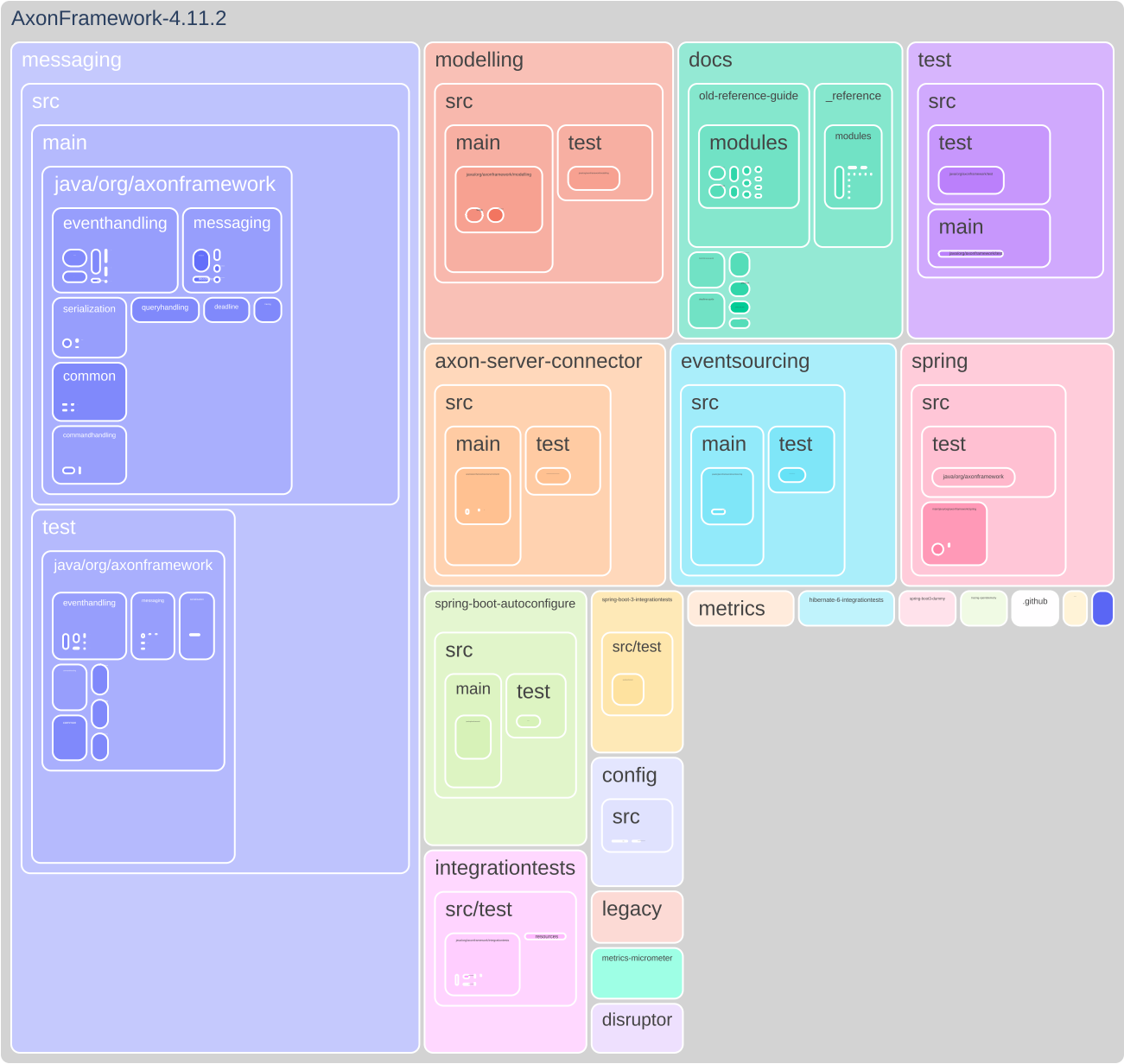
Data Preview

	fileCount	authorCount	commitCount	daysSinceLastCommit	daysSinceLastCreation	daysSinceLastModification
count	563.000000	563.000000	563.000000	563.000000	563.000000	563.000000
mean	24.492007	19.548845	283.804618	224.602131	1091.884547	317.534636
std	117.928146	21.254240	630.075836	137.280170	968.702108	373.632167
min	1.000000	2.000000	3.000000	25.000000	93.000000	24.000000
25%	1.000000	7.000000	27.000000	151.000000	333.000000	150.000000
50%	4.000000	12.000000	80.000000	293.000000	852.000000	292.000000
75%	10.000000	25.000000	262.000000	293.000000	1906.000000	292.000000
max	2250.000000	212.000000	9638.000000	2443.000000	5570.000000	2921.000000

	directoryPath	directoryParentPat
0	AxonFramework-4.11.2/assembly	AxonFramework-4.11.
1	AxonFramework-4.11.2/axon-server-connector/src/main/java/org/axonframework/axonserver/connector/heartbeat/source	AxonFramework-4.11.2/axon-serve connector/src/main/java/org/axonframework/axonserver/connector/heartbea
2	AxonFramework-4.11.2/axon-server-connector/src/main/resources/META-INF/services	AxonFramework-4.11.2/axon-server-connector/src/main/resource
3	AxonFramework-4.11.2/axon-server-connector/src/test/java/org/axonframework/axonserver/connector/event/util	AxonFramework-4.11.2/axon-serve connector/src/test/java/org/axonframework/axonserver/connector/ever
4	AxonFramework-4.11.2/config/src/test/resources/META-INF	AxonFramework-4.11.2/config/src/test/resource
5	AxonFramework-4.11.2/coverage-report	AxonFramework-4.11.
6	AxonFramework-4.11.2/docs/_playbook/localLinks	AxonFramework-4.11.2/docs/_playboc
7	AxonFramework-4.11.2/docs/_reference/modules/ROOT/attachments	AxonFramework-4.11.2/docs/_reference/modules/ROO
8	AxonFramework-4.11.2/docs/_reference/modules/ROOT/partials	AxonFramework-4.11.2/docs/_reference/modules/ROO
9	AxonFramework-4.11.2/docs/_reference/modules/axon-server-connector/pages	AxonFramework-4.11.2/docs/_reference/modules/axon-server-connecto
10	AxonFramework-4.11.2/docs/_reference/modules/axon-server-connector/partials	AxonFramework-4.11.2/docs/_reference/modules/axon-server-connecto
11	AxonFramework-4.11.2/docs/_reference/modules/configuration/pages	AxonFramework-4.11.2/docs/_reference/modules/configuratio
12	AxonFramework-4.11.2/docs/_reference/modules/configuration/partials	AxonFramework-4.11.2/docs/_reference/modules/configuratio
13	AxonFramework-4.11.2/docs/_reference/modules/disruptor/pages	AxonFramework-4.11.2/docs/_reference/modules/disrupto
14	AxonFramework-4.11.2/docs/_reference/modules/disruptor/partials	AxonFramework-4.11.2/docs/_reference/modules/disrupto
15	AxonFramework-4.11.2/docs/_reference/modules/eventsourcing/pages	AxonFramework-4.11.2/docs/_reference/modules/eventsourcin
16	AxonFramework-4.11.2/docs/_reference/modules/eventsourcing/partials	AxonFramework-4.11.2/docs/_reference/modules/eventsourcin
17	AxonFramework-4.11.2/docs/_reference/modules/legacy/pages	AxonFramework-4.11.2/docs/_reference/modules/legac
18	AxonFramework-4.11.2/docs/_reference/modules/legacy/partials	AxonFramework-4.11.2/docs/_reference/modules/legac
19	AxonFramework-4.11.2/docs/_reference/modules/metrics/pages	AxonFramework-4.11.2/docs/_reference/modules/metric
20	AxonFramework-4.11.2/docs/_reference/modules/metrics/partials	AxonFramework-4.11.2/docs/_reference/modules/metric
21	AxonFramework-4.11.2/docs/_reference/modules/micrometer/pages	AxonFramework-4.11.2/docs/_reference/modules/micromete
22	AxonFramework-4.11.2/docs/_reference/modules/micrometer/partials	AxonFramework-4.11.2/docs/_reference/modules/micromete
23	AxonFramework-4.11.2/docs/_reference/modules/modeling/partials	AxonFramework-4.11.2/docs/_reference/modules/modelin
24	AxonFramework-4.11.2/docs/_reference/modules/spring/pages	AxonFramework-4.11.2/docs/_reference/modules/sprin
25	AxonFramework-4.11.2/docs/_reference/modules/spring/partials	AxonFramework-4.11.2/docs/_reference/modules/sprin
26	AxonFramework-4.11.2/docs/_reference/modules/springboot-starter/pages	AxonFramework-4.11.2/docs/_reference/modules/springboot-starte
27	AxonFramework-4.11.2/docs/_reference/modules/springboot-starter/partials	AxonFramework-4.11.2/docs/_reference/modules/springboot-starte
28	AxonFramework-4.11.2/docs/_reference/modules/test/pages	AxonFramework-4.11.2/docs/_reference/modules/tes
29	AxonFramework-4.11.2/docs/_reference/modules/test/partials	AxonFramework-4.11.2/docs/_reference/modules/tes

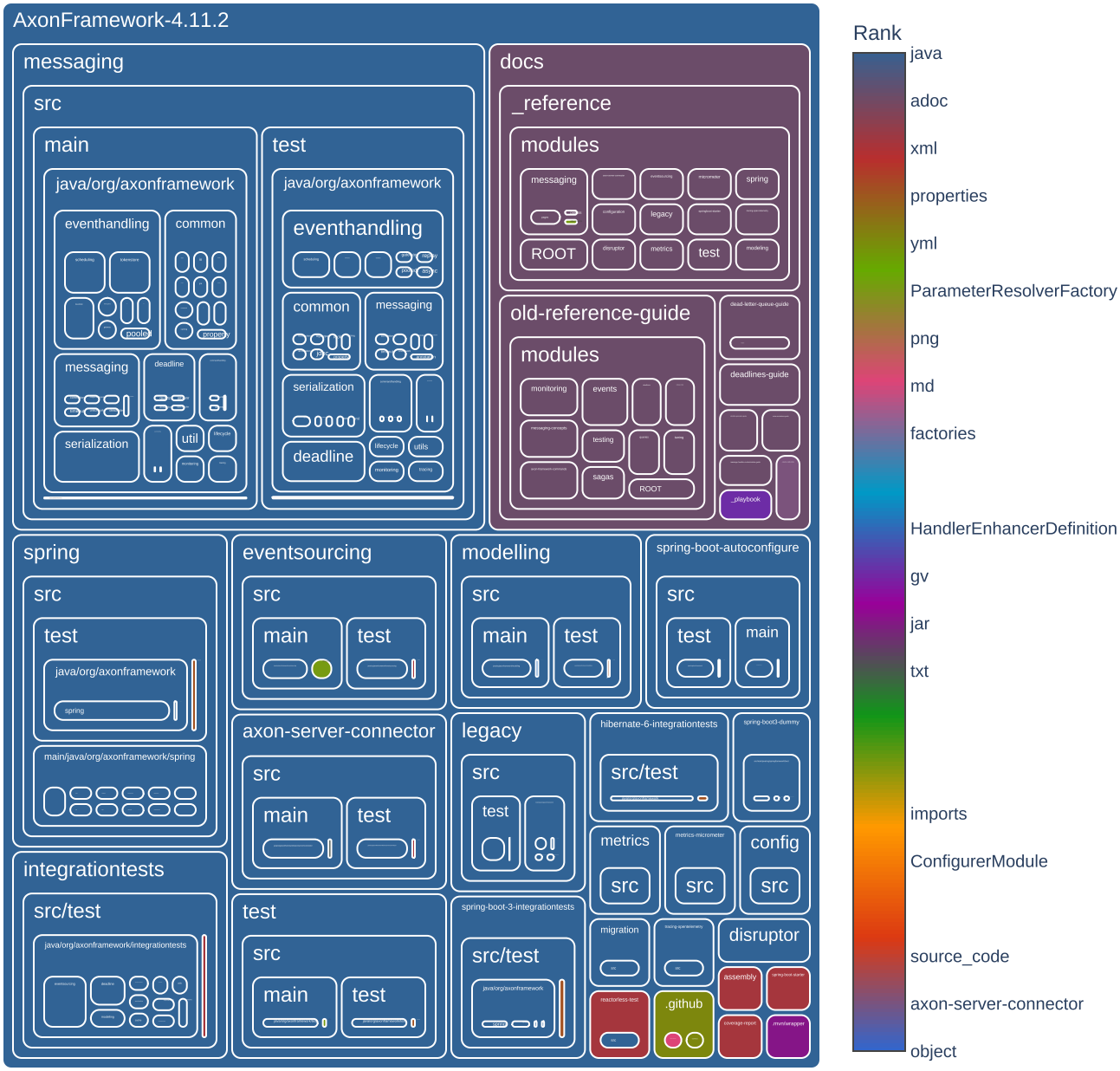
Number of files per directory

Directories and their file count



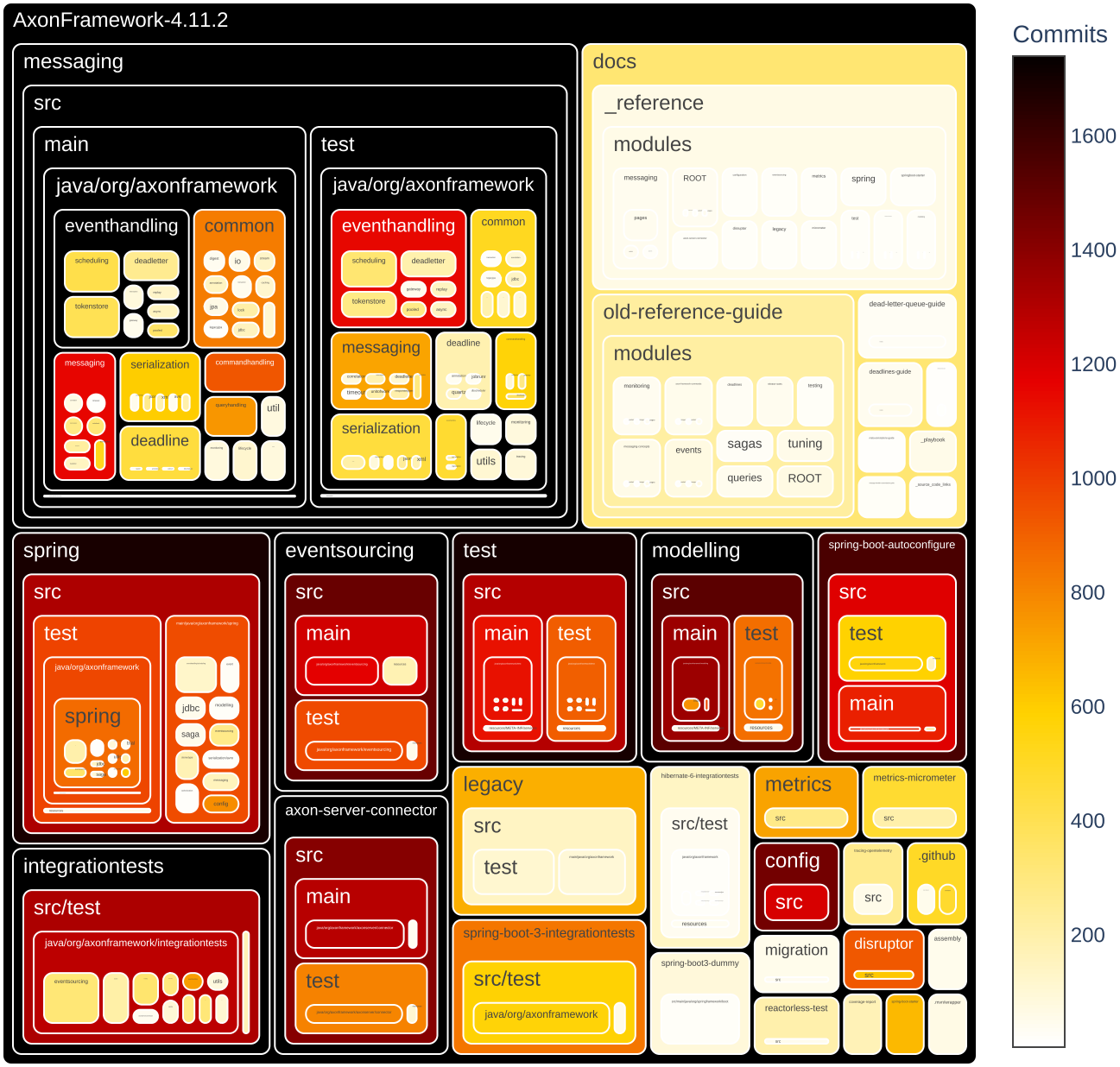
Most frequent file extension per directory

Most frequent file extension per directory



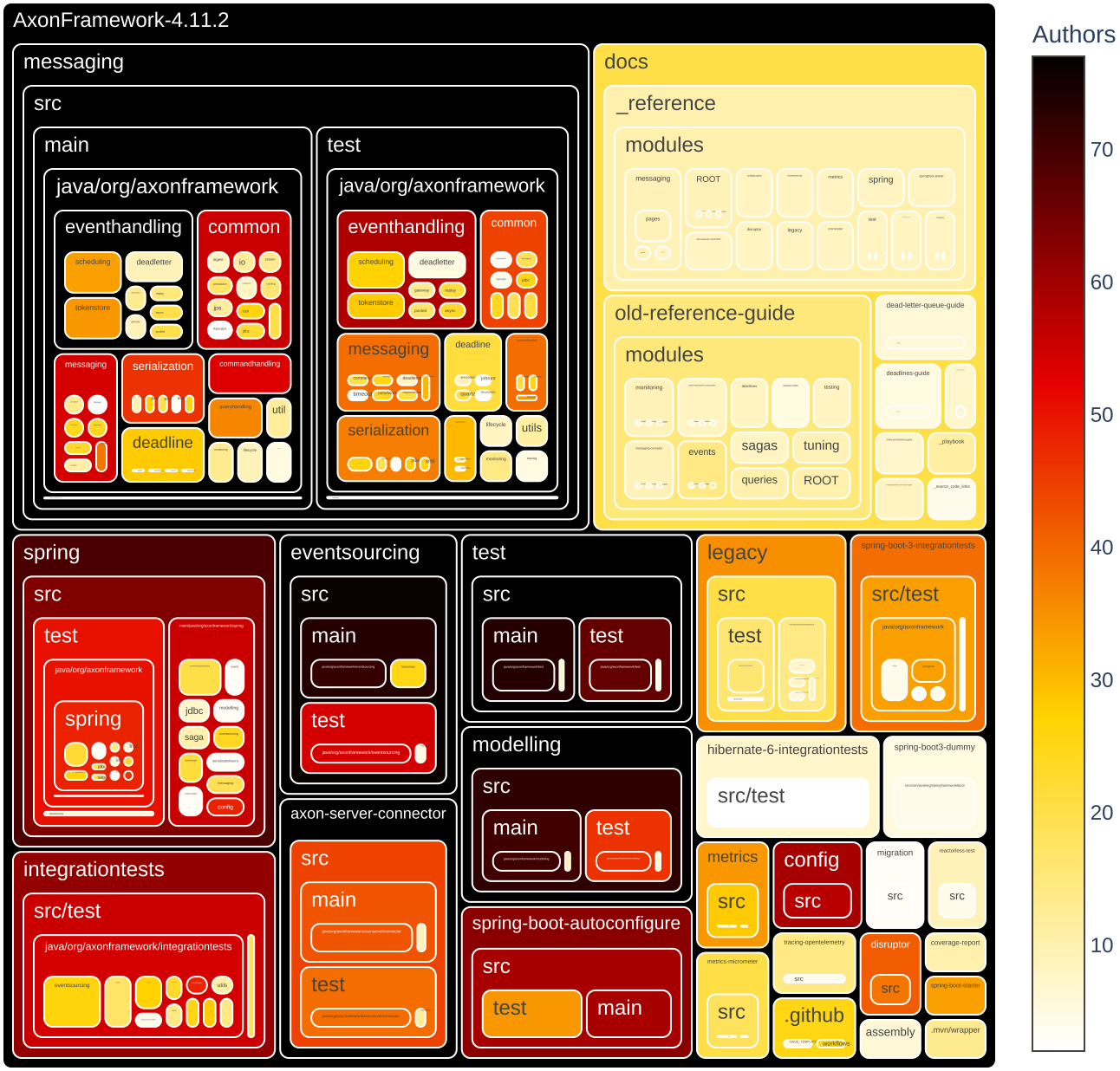
Number of commits per directory

Number of git commits



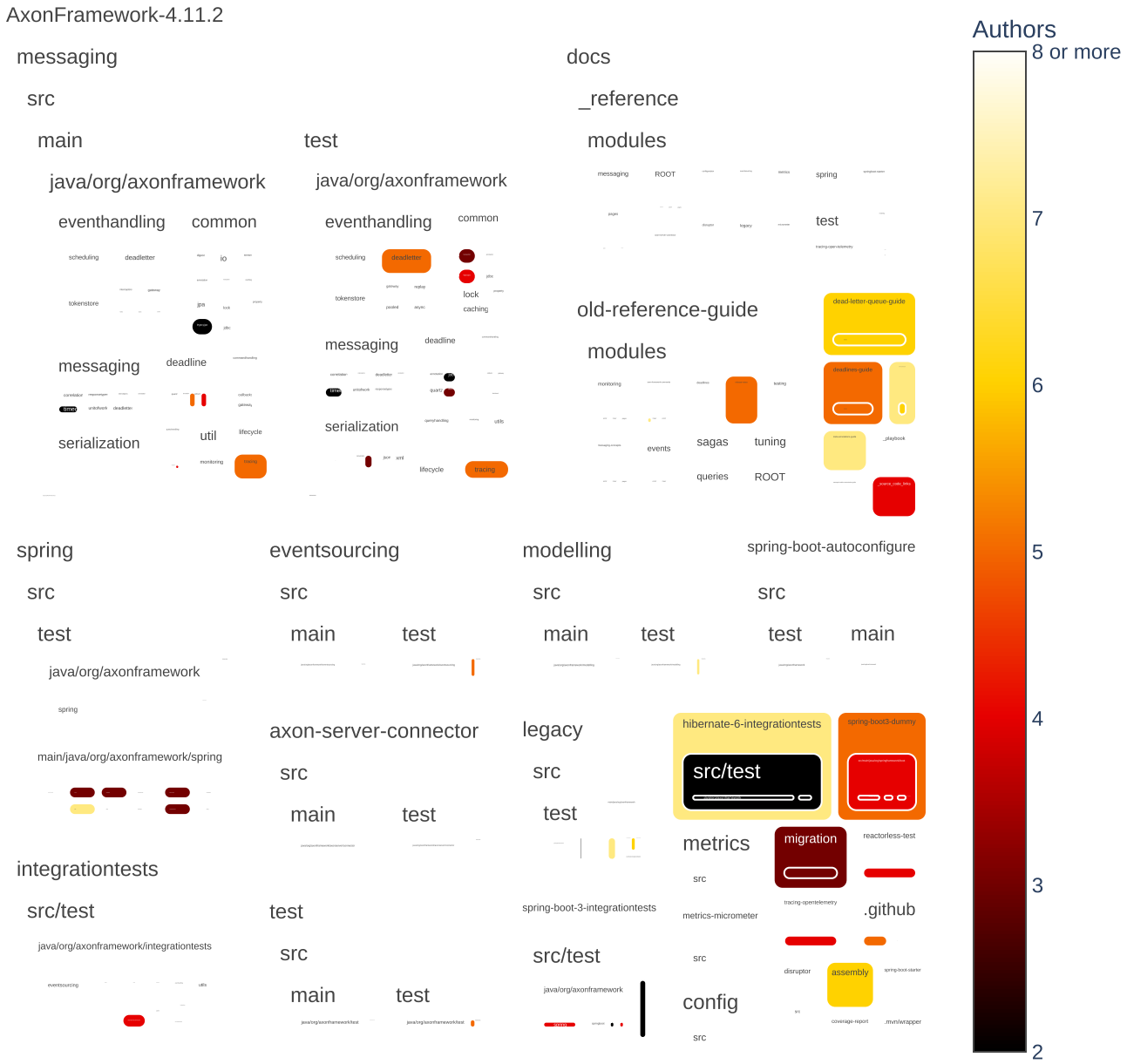
Number of distinct authors per directory

Number of distinct commit authors



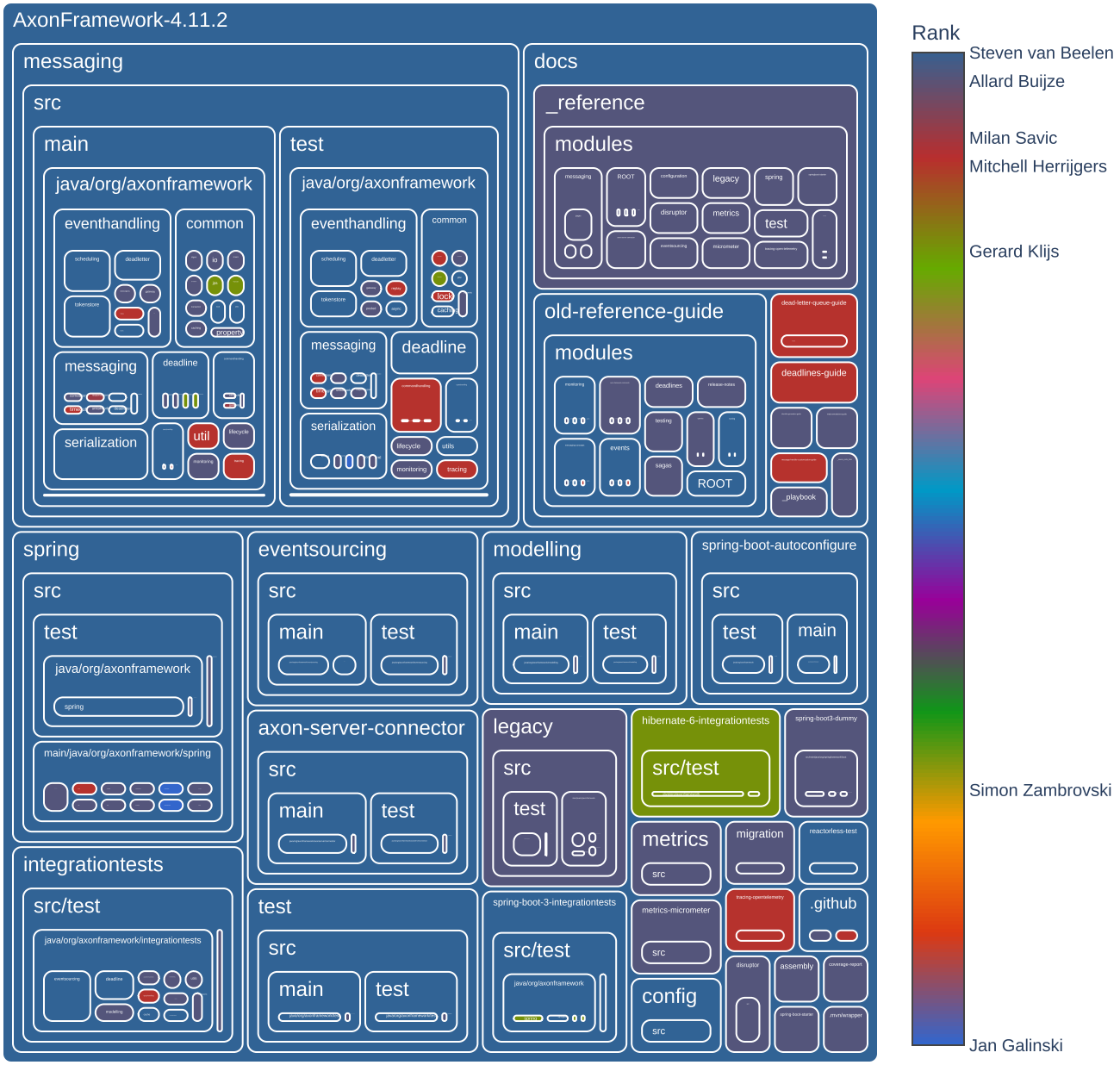
Directories with very few different authors

Number of distinct commit authors (red/black = only one or very few authors)



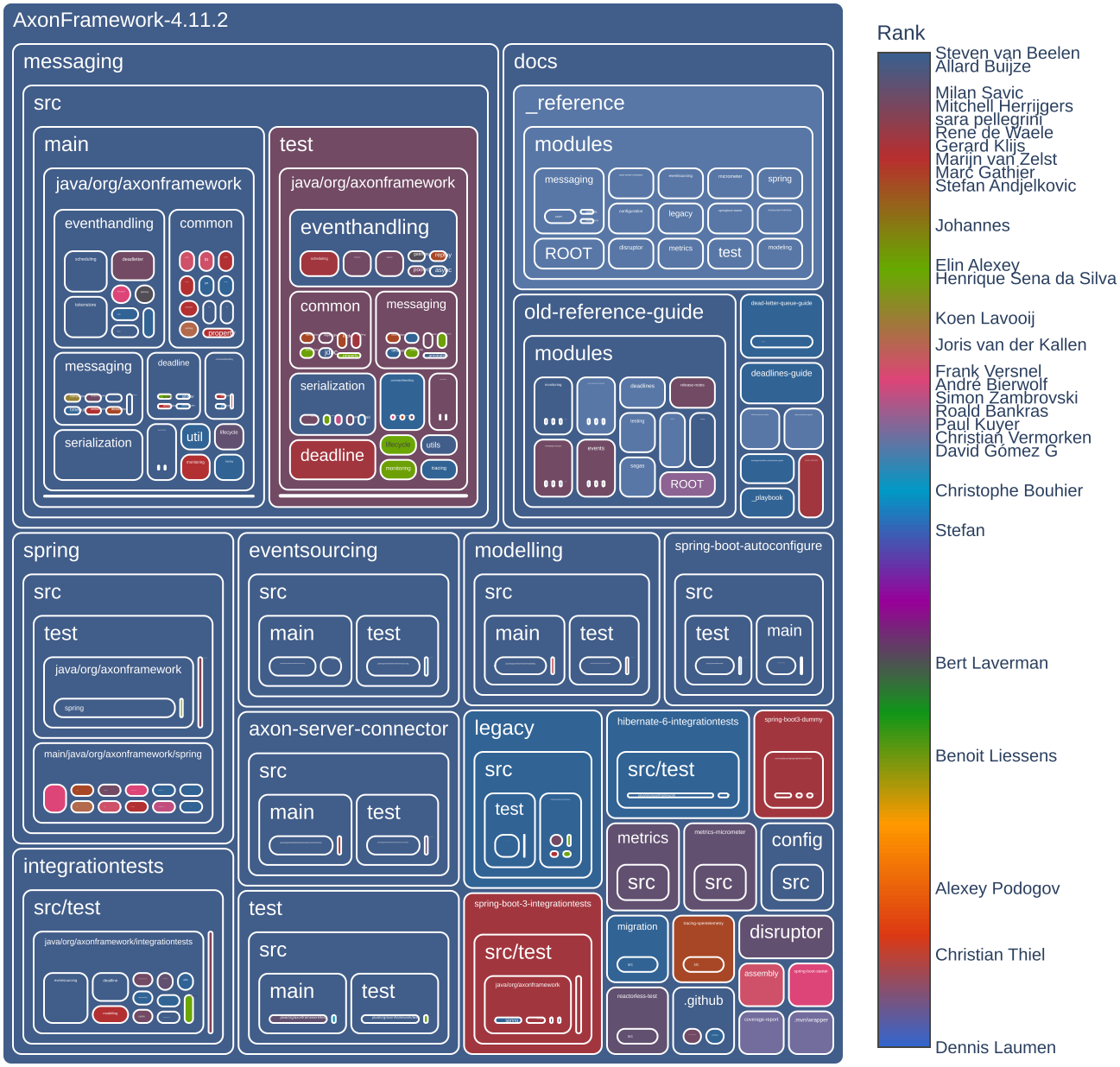
Main author per directory

Main authors with highest number of commits



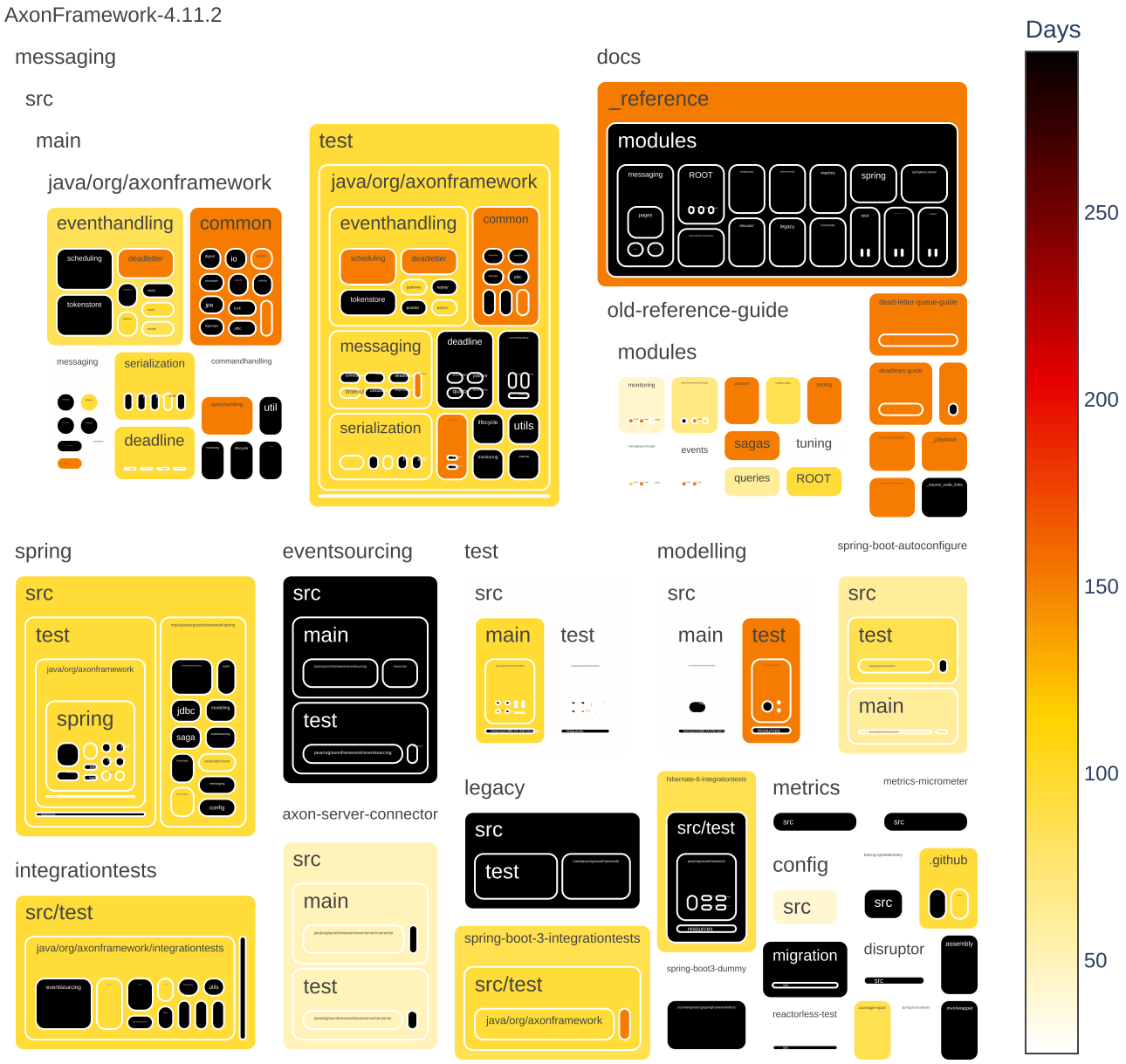
Second author per directory

Second author with the second highest number of commits



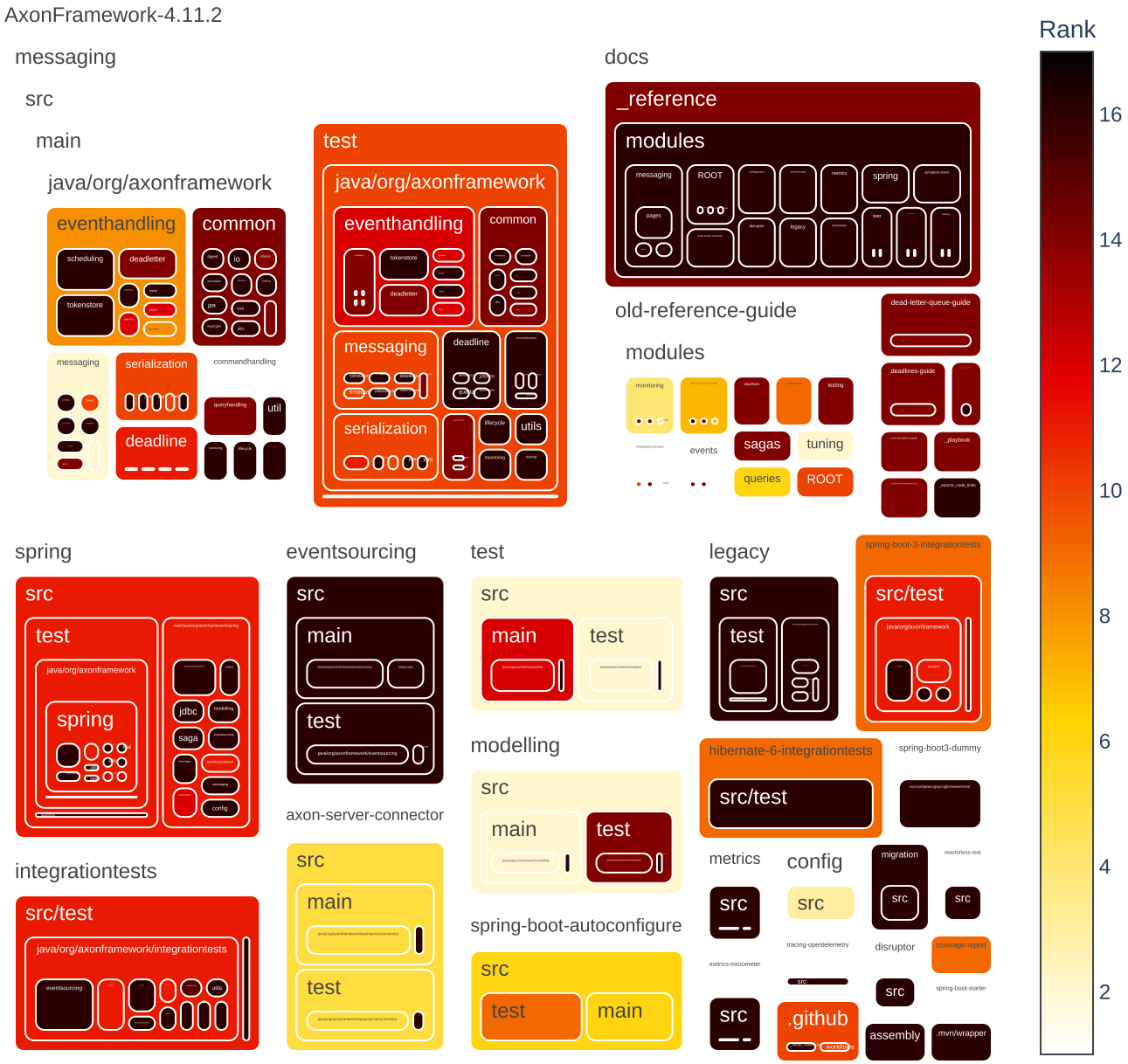
Days since last commit per directory

Days since last commit



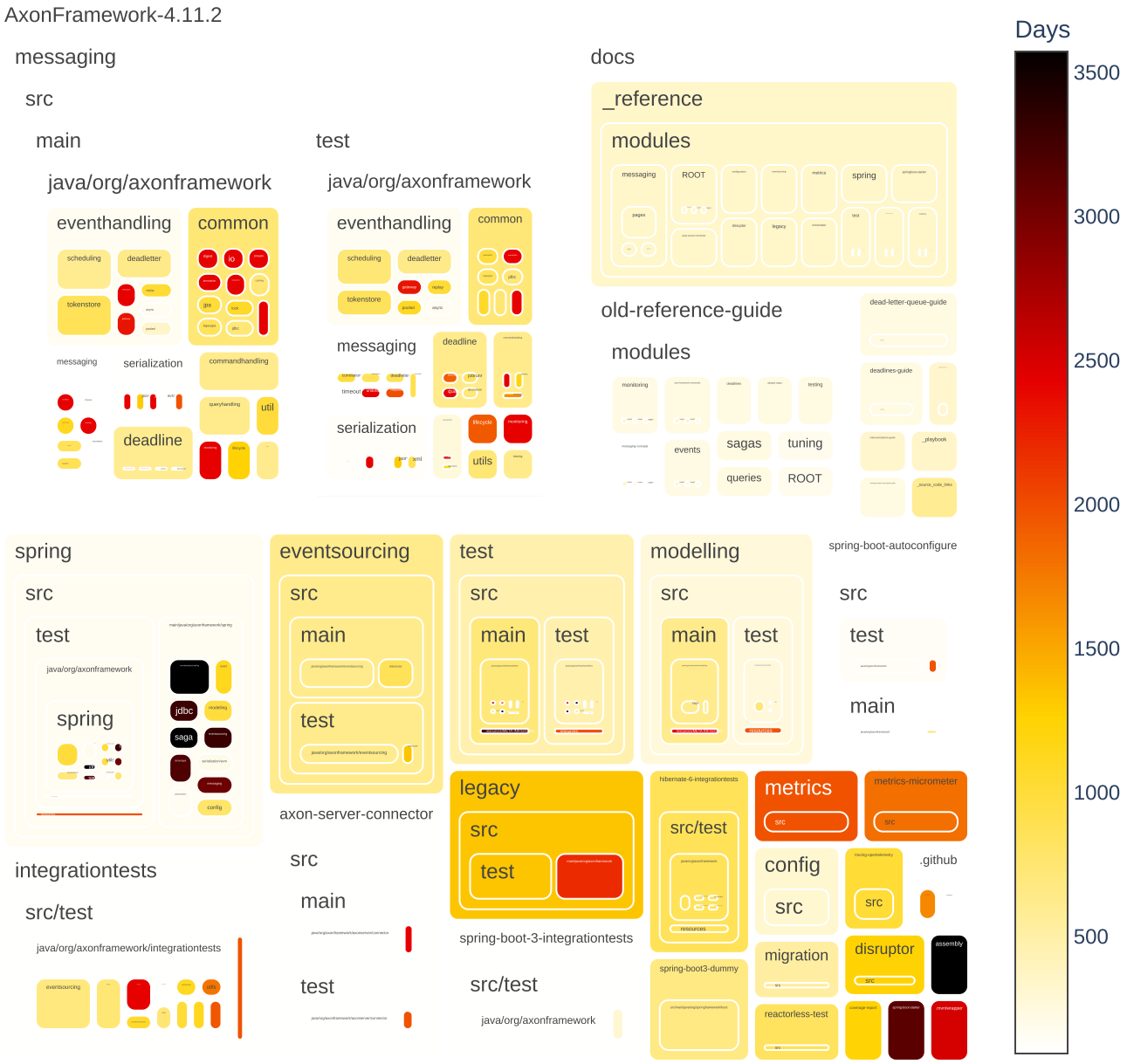
Days since last commit per directory (ranked)

Rank of days since last commit



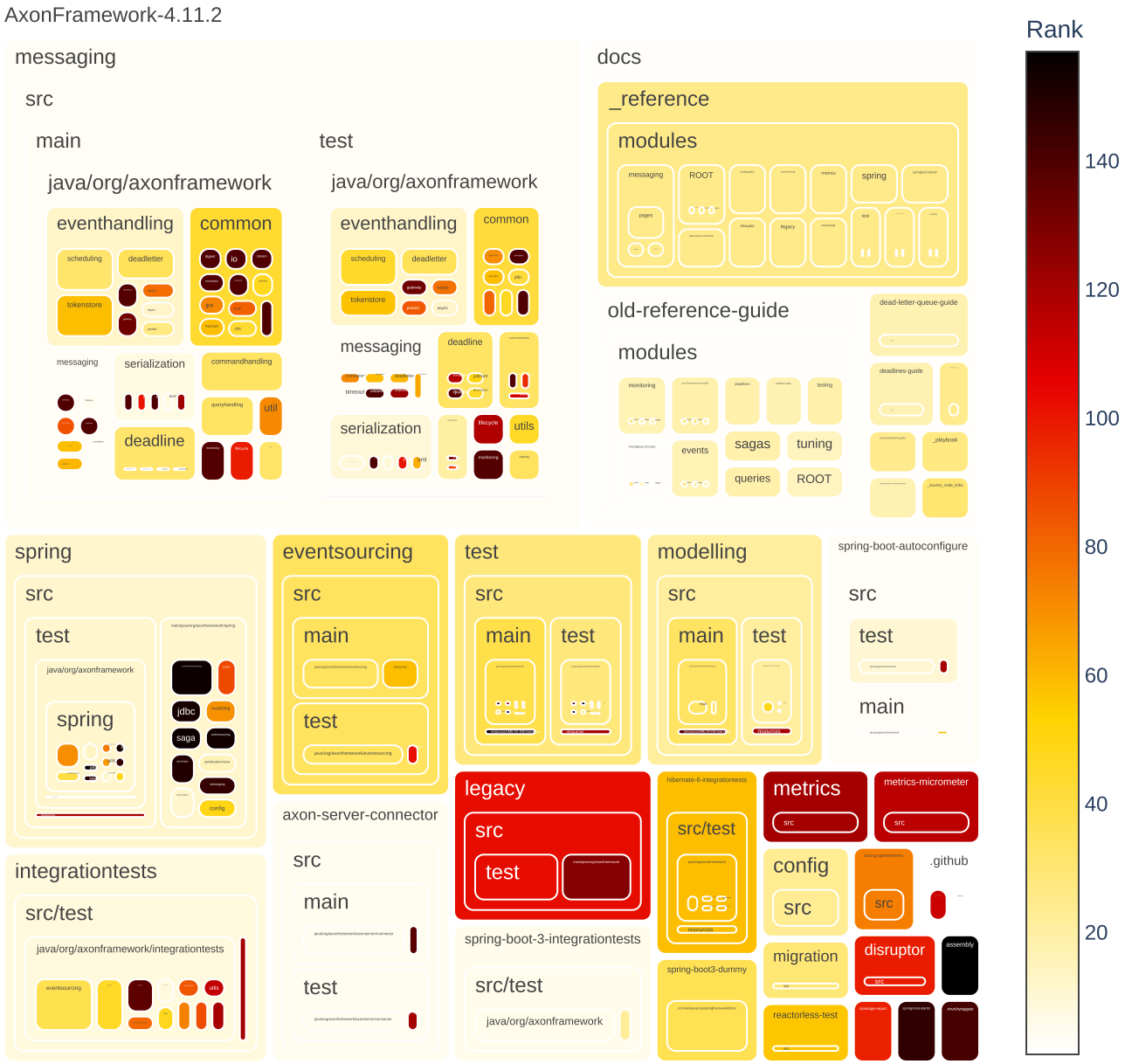
Days since last file creation per directory

Days since last file creation



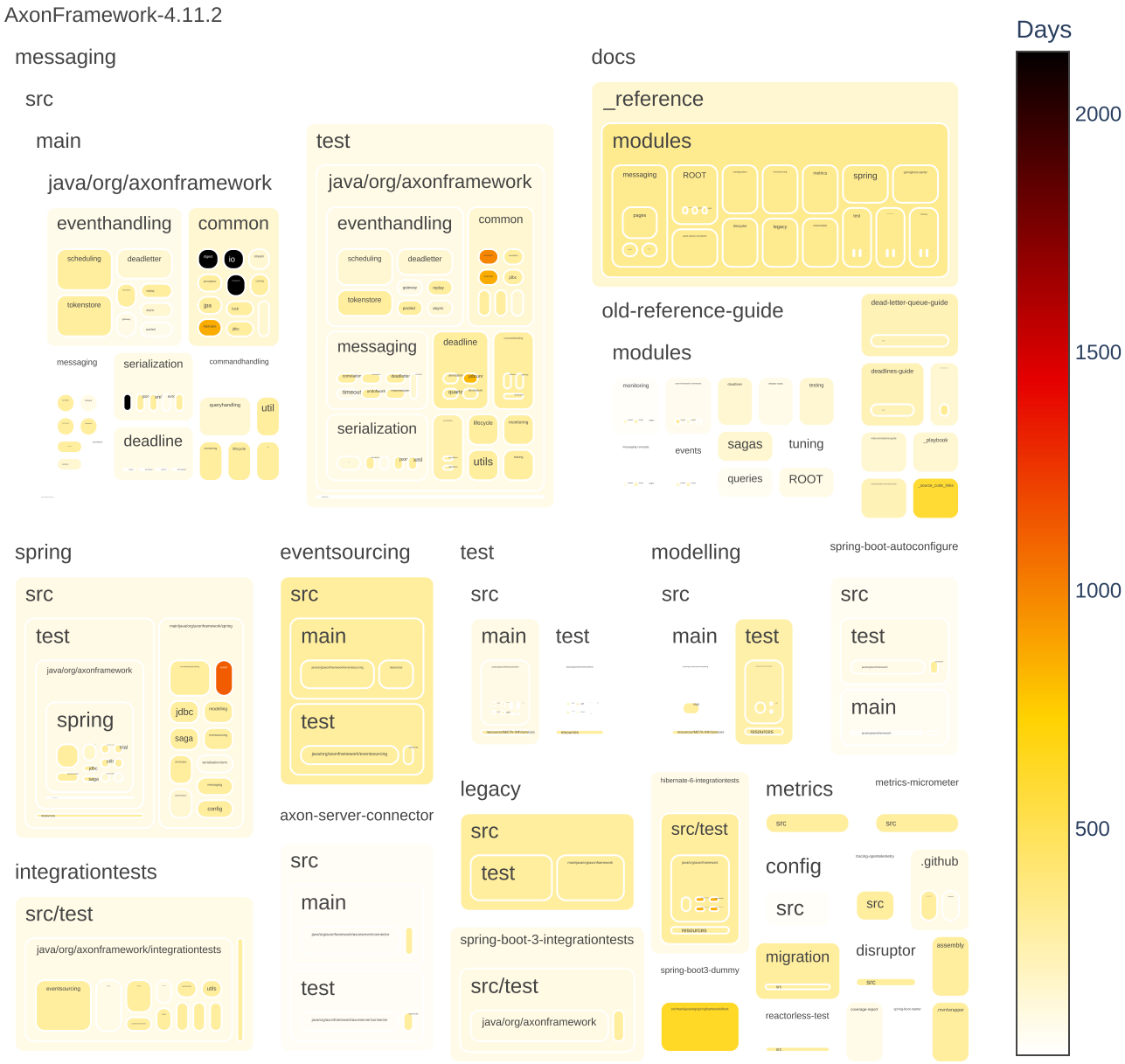
Days since last file creation per directory (ranked)

Rank of days since last file creation



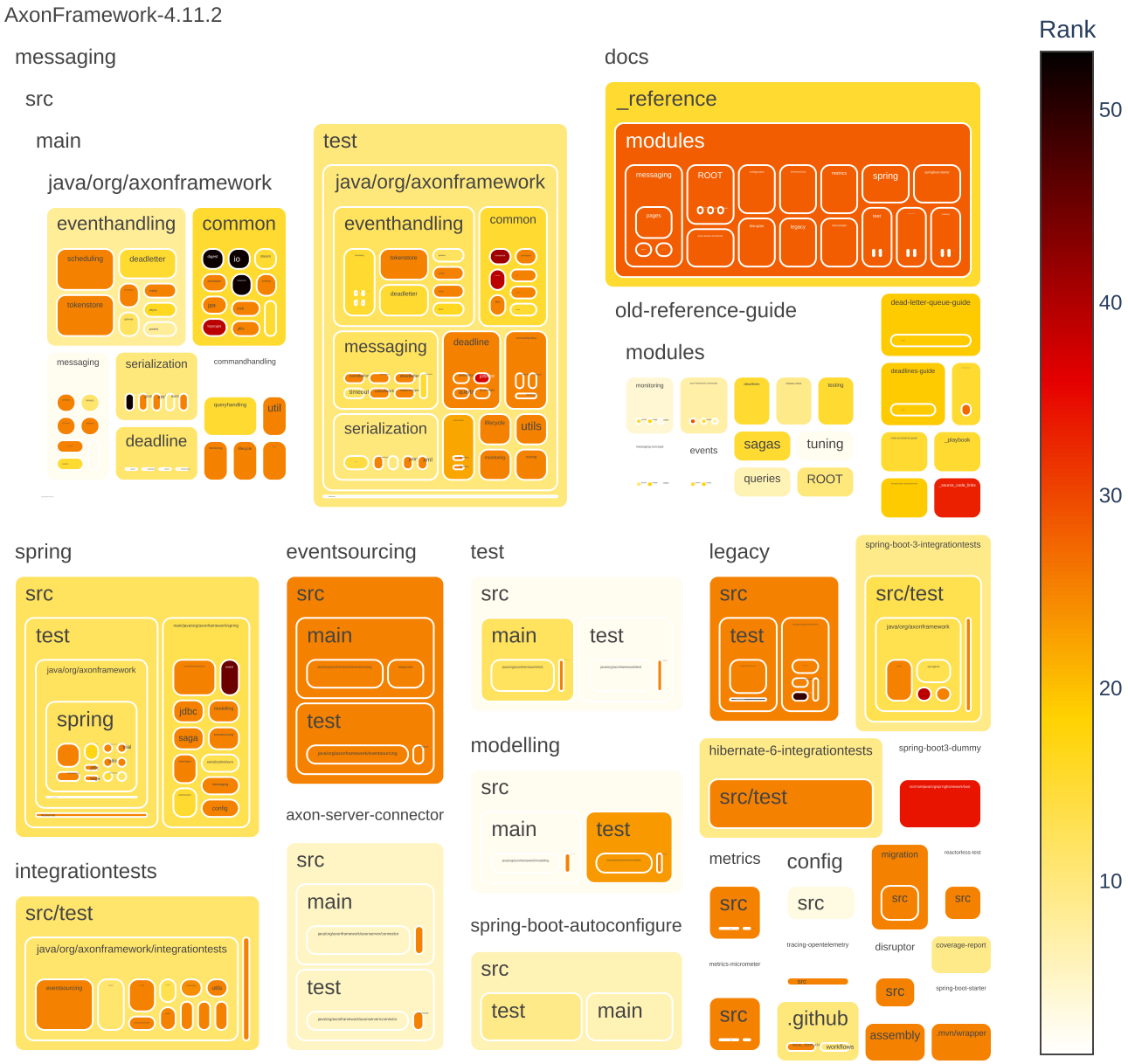
Days since last file modification per directory

Days since last file modification



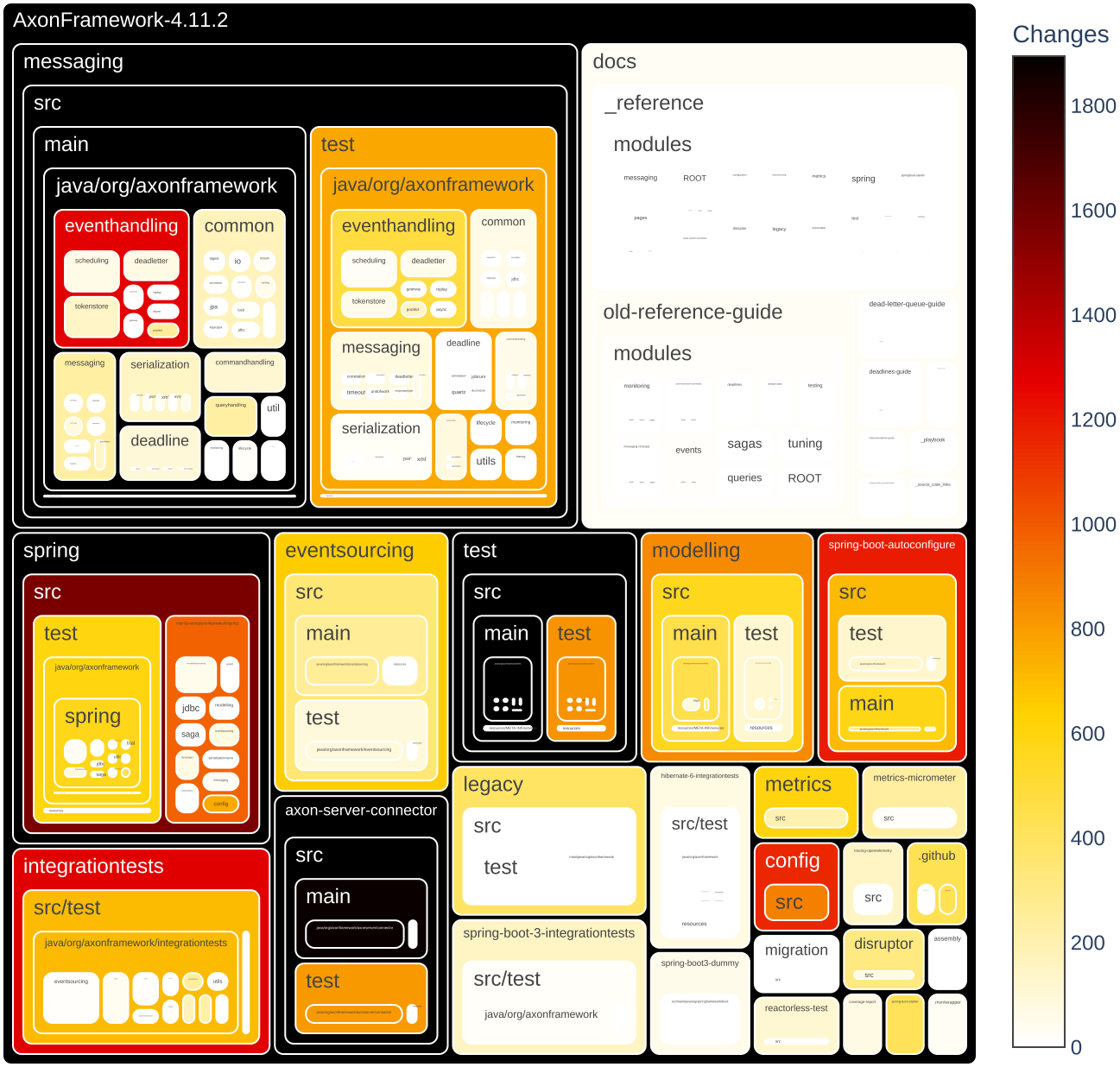
Days since last file modification per directory (ranked)

Rank of days since last file modification



File changed frequently with other files

Pairwise file changes



Filecount per commit

Shows how many commits had changed one file, how many had changed two files, and so on. The chart is limited to 30 lines for improved readability. The data preview also includes overall statistics including the number of commits that are filtered out in the chart.

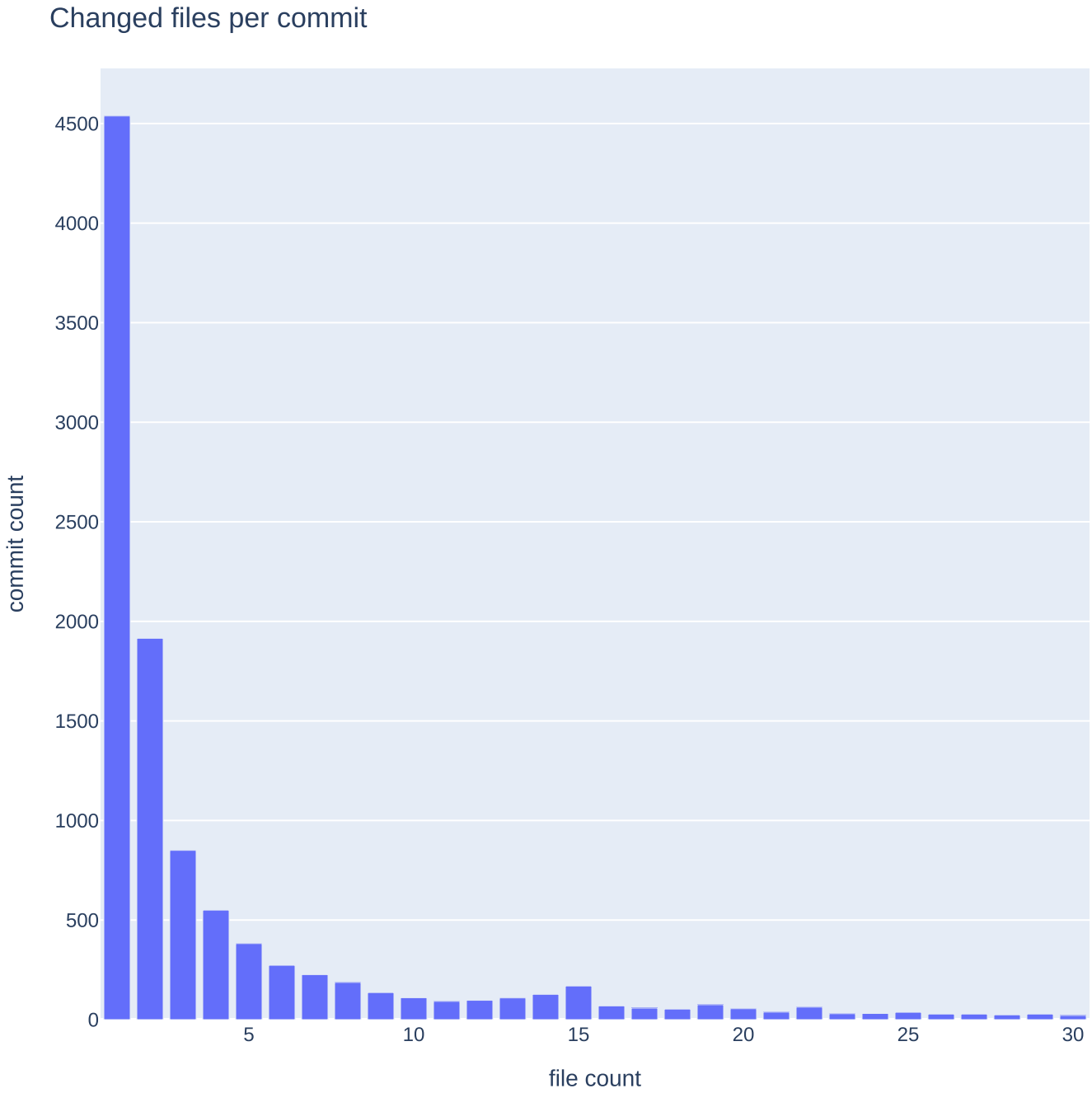
Preview data

```
Sum of commits that changed more than 30 files (each) = 961
Max changed files with one commit = 2178
```

	filesPerCommit	commitCount
count	333.000000	333.000000
mean	344.075075	34.114114
std	440.375494	276.321358
min	1.000000	1.000000
25%	84.000000	1.000000
50%	186.000000	2.000000
75%	428.000000	5.000000
max	2178.000000	4538.000000

	filesPerCommit	commitCount
0	1	4538
1	2	1915
2	3	850
3	4	550
4	5	382
5	6	273
6	7	226
7	8	187
8	9	136
9	10	110
10	11	92
11	12	97
12	13	109
13	14	127
14	15	169
15	16	69
16	17	58
17	18	53
18	19	75
19	20	55
20	21	38
21	22	63
22	23	30
23	24	31
24	25	37
25	26	28
26	27	28
27	28	24
28	29	28
29	30	21

Bar chart with the number of files per commit distribution



Pairwise Changed Files vs. Dependency Weight

This section explores the correlation between how often pairs of files are changed together (common commit count) and their dependency weight. Note that these results should be interpreted cautiously, as comparing pairwise changes and dependencies is inherently challenging.

Considerations

- **Historical vs. Current State**: Pairwise changes reflect the entire git history, while dependency weight represents the current state of the codebase.

- **Commit Granularity**: Developers may use different commit strategies, such as squashing changes into a single commit or creating fine-grained commits. Ideally, each commit should represent a single semantic change for accurate analysis.
- **Dependency Representation**: Some file types (e.g., Java files with import statements) clearly define dependencies, while others (e.g., shell scripts, XML, YAML) lack explicit dependency relationships.
- **Repository Characteristics**: Repositories with generated code may have many large commits, while stabilized repositories may only update configuration files for dependency changes.

Data Preview

```
Received notification from DBMS server: {severity: WARNING} {code: Neo.ClientNotifica
tion.Statement.UnknownPropertyKeyWarning} {category: UNRECOGNIZED} {title: The provid
ed property key is not in the database} {description: One of the property names in yo
ur query is not available in the database, make sure you didn't misspell it or that t
he label is available when you run this statement in your application (the missing pr
operty name is: cardinality)} {position: line: 9, column: 48, offset: 576} for query:
"// List pair of files that were changed together and that have a declared dependency
between each other.\n \n MATCH (firstCodeFile:File)-[dependency:DEPENDS_ON]->(secondC
odeFile:File)\n MATCH (firstCodeFile)-[pairwiseChange:CHANGED_TOGETHER_WITH]-(secondC
odeFile)\n //De-duplicating the pairs of files isn't necessary, because the dependenc
y relation is directed.\n //WHERE elementId(firstCodeFile) < elementId(secondCodeFil
e)\n  WITH  firstCodeFile.fileName      AS firstFileName\n       ,secondCodeFile.file
Name     AS secondFileName\n       ,coalesce(dependency.weight, dependency.cardinalit
y)    AS dependencyWeight\n       ,pairwiseChange.commitCount  AS commitCount\n
,dependency.fileDistanceAsFewestChangeDirectoryCommands AS fileDistanceAsFewestChange
DirectoryCommands\n RETURN dependencyWeight\n       ,commitCount\n       ,fileDistanc
eAsFewestChangeDirectoryCommands\n       // ,count(*)                    AS occurrenc
es\n       // ,collect(firstFileName + ' -> ' + secondFileName)[0..3] AS examples\n O
RDER BY dependencyWeight, commitCount\n \n // MATCH (firstCodeFile:File)-[dependency:
DEPENDS_ON]->(secondCodeFile:File)\n // MATCH (firstCodeFile)-[pairwiseChange:CHANGED
_TOGETHER_WITH]-(secondCodeFile)\n // WHERE elementId(firstCodeFile) < elementId(seco
ndCodeFile)\n // RETURN firstCodeFile.fileName  AS firstFileName\n //       ,secondCo
deFile.fileName AS secondFileName\n //       ,dependency.weight           AS dependen
cyWeight\n //       ,pairwiseChange.commitCount  AS commitCount\n // ORDER BY depende
ncyWeight, commitCount\n \n //  MATCH (g1:!Git&File)-[relation:CHANGED_TOGETHER_WITH|
DEPENDS_ON]-(g2:!Git&File) \n //   WITH count(DISTINCT relation)   AS relatedFilesCou
nt\n //       ,collect(DISTINCT relation) AS relations\n // UNWIND relations AS relat
ion\n //   WITH relatedFilesCount\n //       ,coalesce(relation.commitCount, 0)
AS commitCount\n //       ,coalesce(relation.weight, 0)
AS dependencyWeight\n //       ,coalesce(relation.fileDistanceAsFewestChangeDirectory
Commands, 0) AS fileDistanceAsFewestChangeDirectoryCommands\n // RETURN dependencyWei
ght\n //       ,commitCount\n //       ,fileDistanceAsFewestChangeDirectoryCommands\n
// ORDER BY dependencyWeight, commitCount\n"
```

	dependencyWeight	commitCount	fileDistanceAsFewestChangeDirectoryCommands
0	1	12	0
1	1	12	0
2	1	12	0
3	1	12	2
4	1	12	0
5	1	12	4
6	1	12	4
7	1	12	0
8	1	12	0
9	1	12	4
10	1	13	0
11	1	13	4
12	1	13	0
13	1	13	4
14	1	14	2
15	1	14	4
16	1	15	2
17	1	15	0
18	1	15	0
19	1	16	0

Data Statistics

'Pairwise changed git files compared to dependency weights - Overall statistics'

	dependencyWeight	commitCount	fileDistanceAsFewestChangeDirectoryCommands
count	166.000000	166.000000	166.000000
mean	9.566265	25.253012	1.373494
std	19.257863	16.677579	1.570260
min	1.000000	12.000000	0.000000
25%	1.250000	14.250000	0.000000
50%	3.000000	19.000000	0.000000
75%	10.000000	28.000000	2.000000
max	154.000000	95.000000	4.000000

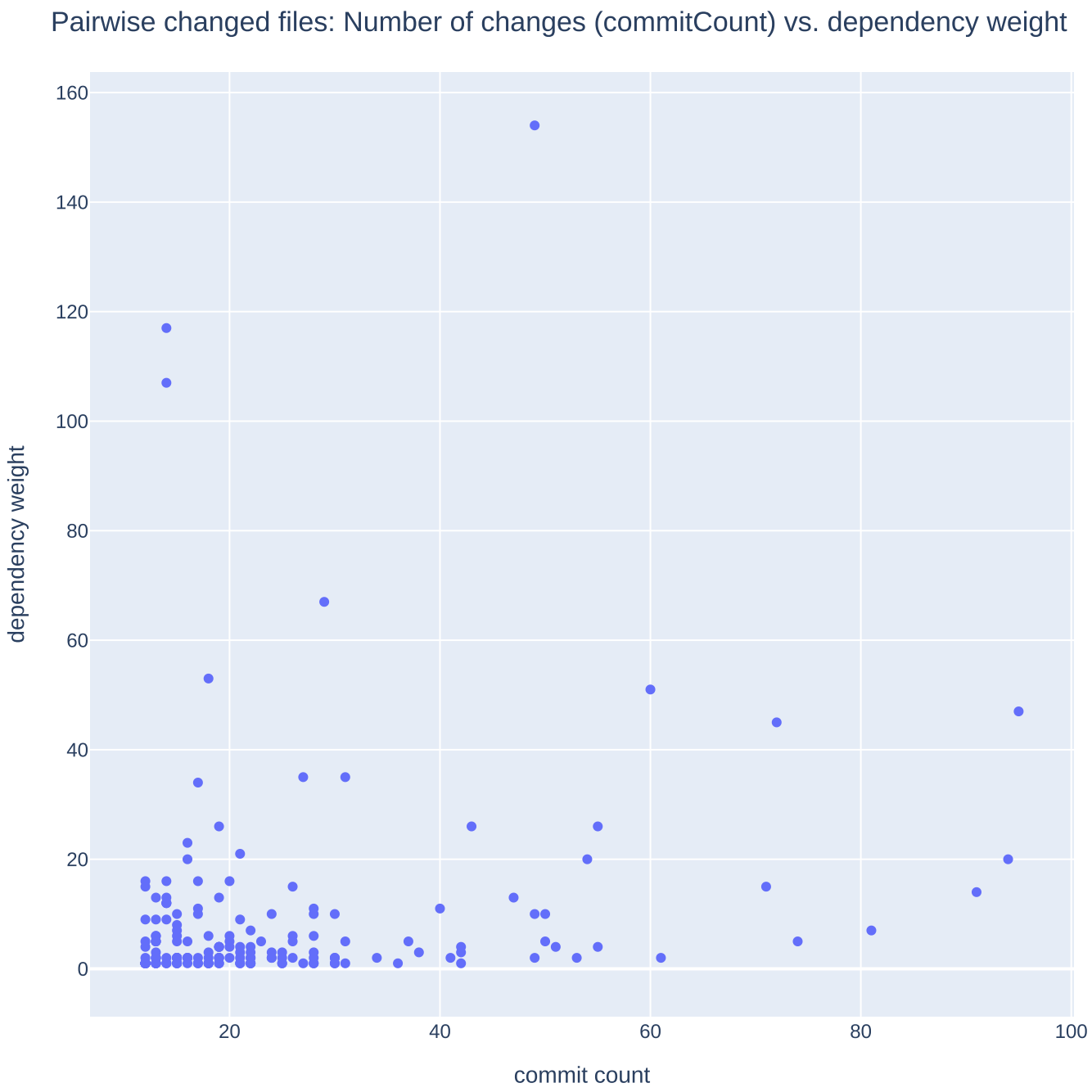
'Pairwise changed git files compared to dependency weights - Pearson Correlation'

	dependencyWeight	commitCount	fileDistanceAsFewestChangeDirectoryCommands
dependencyWeight	1.000000	0.204462	-0.143721
commitCount	0.204462	1.000000	-0.143874
fileDistanceAsFewestChangeDirectoryCommands	-0.143721	-0.143874	1.000000

'Pairwise changed git files compared to dependency weights - Spearman Correlation'

	dependencyWeight	commitCount	fileDistanceAsFewestChangeDirectoryCommands
dependencyWeight	1.000000	0.212543	-0.089280
commitCount	0.212543	1.000000	-0.132659
fileDistanceAsFewestChangeDirectoryCommands	-0.089280	-0.132659	1.000000

```
'Pearson Correlation with p-value for commitCount and dependencyWeight'
PearsonRResult(statistic=0.20446197739019414, pvalue=0.008232632527754543)
'Spearman Correlation with p-value for commitCount and dependencyWeight'
SignificanceResult(statistic=0.21254323842799402, pvalue=0.005973735857352672)
```



WordCloud of git authors

	word	frequency
0	Steven van Beelen	3402
1	Allard Buijze	2722
2	smcvb	822
3	Milan Savic	449
4	Mitchell Herrijgers	313
5	Gerard Klijs	246
6	Marc Gathier	231
7	sara pellegrini	163
8	Rene de Waele	121
9	Jettro Coenradie	117

Wordcloud of git authors