

CODE HISTORY ANALYSIS AND VISUALISATION

dmitry.kandalov@gmail.com

<http://bit.ly/22RCf12>





This repository

Search

Pull requests Issues Gist



michaelfeathers / delta-flora

Watch 11

★ Unstar 48

Fork 11

Code Issues 0 Pull requests 3 Wiki Pulse Graphs

No description or website provided.

64 commits

1 branch

0 releases

1 contributor

Branch: master

New pull request

New file

Upload files

Find file

HTTPS

https://github.com/mich



Download ZIP

michaelfeathers added smooth and sharpen along with edge_extend

Latest commit 756ae53 on Jun 24, 2015

lib	added smooth and sharpen along with edge_extend	10 months ago
spec	factored all array extensions from analytics.rb	a year ago
.gitignore	added adds per hour	2 years ago
README.rdoc	added example output	a year ago
Rakefile	added spec for CodeEvent	a year ago

README.rdoc

delta-flora

- github.com/michaelfeathers/delta-flora/

```
irb(main):001:0> load 'repository.rb'
```

```
=> true
```

```
irb(main):002:0> load 'array_ext.rb'
```

```
=> true
```

```
irb(main):003:0> events = Repository.new('..').events
```

```
15 +0100 lib/code_event.rb CodeEvent::method_score: 2 24 26, bb681266b6498921008c2fc4ef0191686411f160 added 2015-04-23 09:24:15 +0100
lib/code_event.rb CodeEvent::method_complexity: 2 28 30, bb681266b6498921008c2fc4ef0191686411f160 added 2015-04-23 09:24:15 +0100 lib/
code_event.rb CodeEvent::method_functional_score: 2 32 34, bb681266b6498921008c2fc4ef0191686411f160 added 2015-04-23 09:24:15 +0100 li
b/code_event.rb CodeEvent::git_local: 2 60 62, bb681266b6498921008c2fc4ef0191686411f160 changed 2015-04-23 09:24:15 +0100 lib/reposito
ry.rb Repository::new_event: 11 103 114, bb681266b6498921008c2fc4ef0191686411f160 changed 2015-04-23 09:24:15 +0100 lib/repository.rb
Repository::write: 2 116 118, 7dfde62fbdcf7c2b1c217d76556f43c1ddb77d58 changed 2015-04-25 00:08:35 +0100 lib/code_event.rb CodeEvent::
method_score: 2 24 26, 5690e4ebfb5d7ccdc1f235e79afb8df8443f5994 add
ed 2015-04-25 00:19:43 +0100 lib/analytics.rb method_shark: 3 10 13
]
irb(main):004:0> events.first
=> f75d684dbfcc8e02a07ea49fe7d266fbf51a3ca6 added 2012-12-03 10:58:
26 +0000 analytics.rb month_from_date: 2 5 7
irb(main):005:0> █
```

```
irb(main):004:0> events.first
=> f75d684dbfcc8e02a07ea49fe7d266fbf51a3ca6 added 2012-12-03 10:58:
26 +0000 analytics.rb month_from_date: 2 5 7
irb(main):005:0>
irb(main):006:0*
irb(main):007:0*
irb(main):008:0*
irb(main):009:0* def class_method_count_freq es
irb(main):010:1>   es.group_by(&:class_name)
irb(main):011:1>     .map { |_, v| v.map(&:method_name).uniq.count }
irb(main):012:1>     .freq
irb(main):013:1> end
=> :class_method_count_freq
irb(main):014:0>
irb(main):015:0*
irb(main):016:0*
irb(main):017:0*
irb(main):018:0*
irb(main):019:0*
```

```
irb(main):008:0*
irb(main):009:0* def class_method_count_freq es
irb(main):010:1>   es.group_by(&:class_name)
irb(main):011:1>   .map { |_, v| v.map(&:method_name).uniq.count }
irb(main):012:1>   .freq
irb(main):013:1> end
=> :class_method_count_freq
irb(main):014:0>
irb(main):015:0*
irb(main):016:0*
irb(main):017:0*
irb(main):018:0*
irb(main):019:0* class_method_count_freq(events)
=> [[1, 2], [2, 3], [4, 3], [5, 1], [7, 1], [9, 2], [11, 2], [17, 1],
], [93, 1]]
irb(main):020:0>
irb(main):021:0*
irb(main):022:0*
irb(main):023:0*
```


WHY?

SOFTWARE DESIGN

Your Code As a Crime Scene

Use Forensic Techniques
to Arrest Defects, Bottlenecks, and
Bad Design in Your Programs



PROJECT HISTORY

Predictive Models of Development Teams and the Systems they Build

- [Robert Smallshire](#)

It's awkward to perform science experiments on developers, so let's simulate them instead! The emerging field of software process dynamics applies systems thinking and simulated experiments of software development teams and the systems they build, to inform decisions on projects, process and architecture.

In 1968 Melvin Conway pointed out a seemingly inevitable symmetry between organisations and the software systems they construct. Organisations today are more fluid than 40 years ago, with short developer tenure, and frequent migration of individuals between projects and employers. In this slot we'll examine – and perhaps collect – data on the tenure and productivity of programmers and use this to gain insight into codebases, by simulating their growth with simple stochastic models. From such models, we can make important predictions about the maintainability and long-term viability of software systems, with implications for how we approach software design, documentation and how we assemble teams.

In the second part of the talk we'll build a software dynamics model simulating the growth of a system, and the flow of code change, before concluding with the remarks on the nature of modelling and the basics of making your own models.

GENERIC INSIGHTS

WHY VISUALISE?

A 3-D View of a Chart That Predicts The Economic Future: The Yield Curve

By GREGOR AISCH and AMANDA COX MARCH 18, 2015

● ○ ○ ○ ○ ○ ○ ○ ○ ○

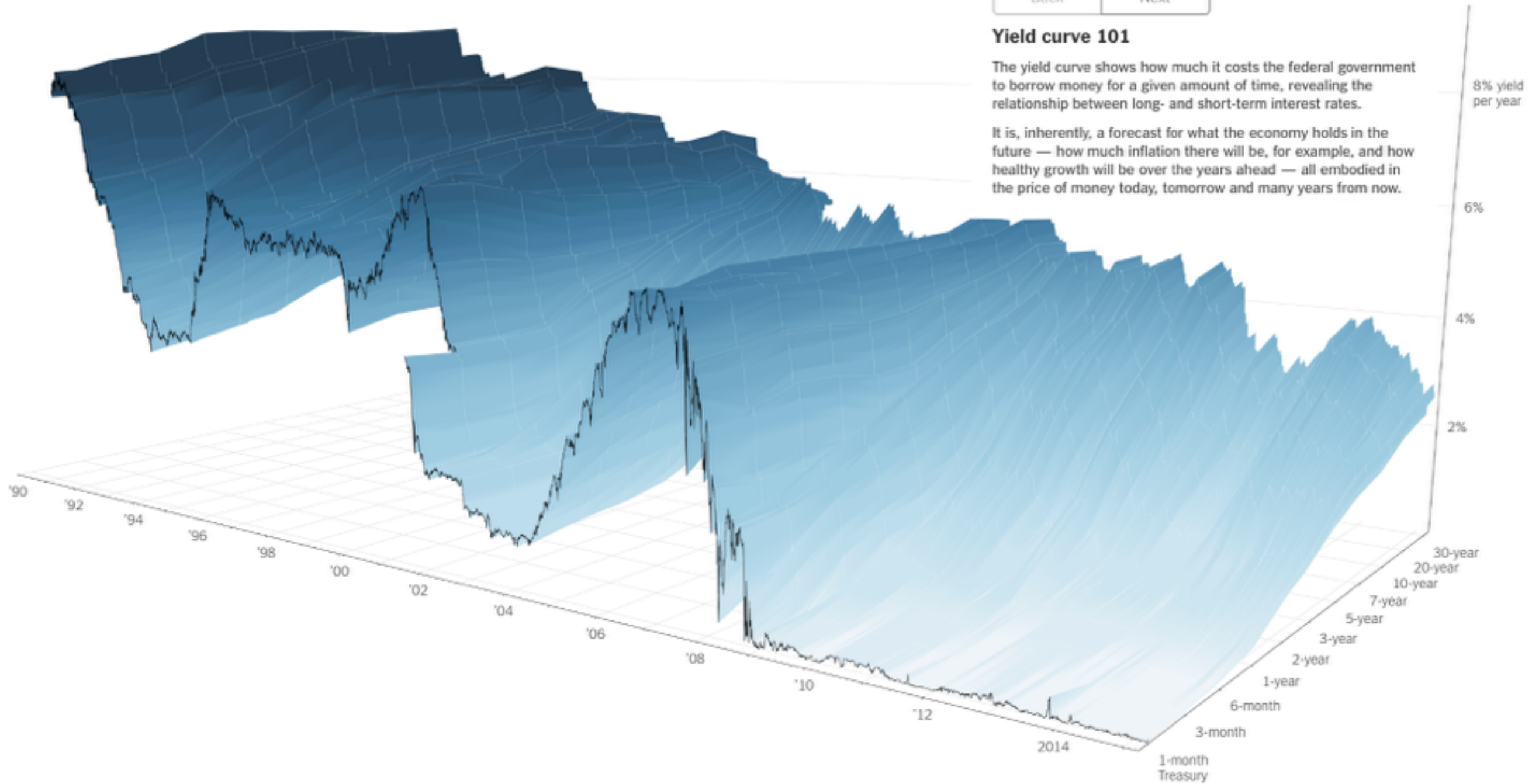
Back

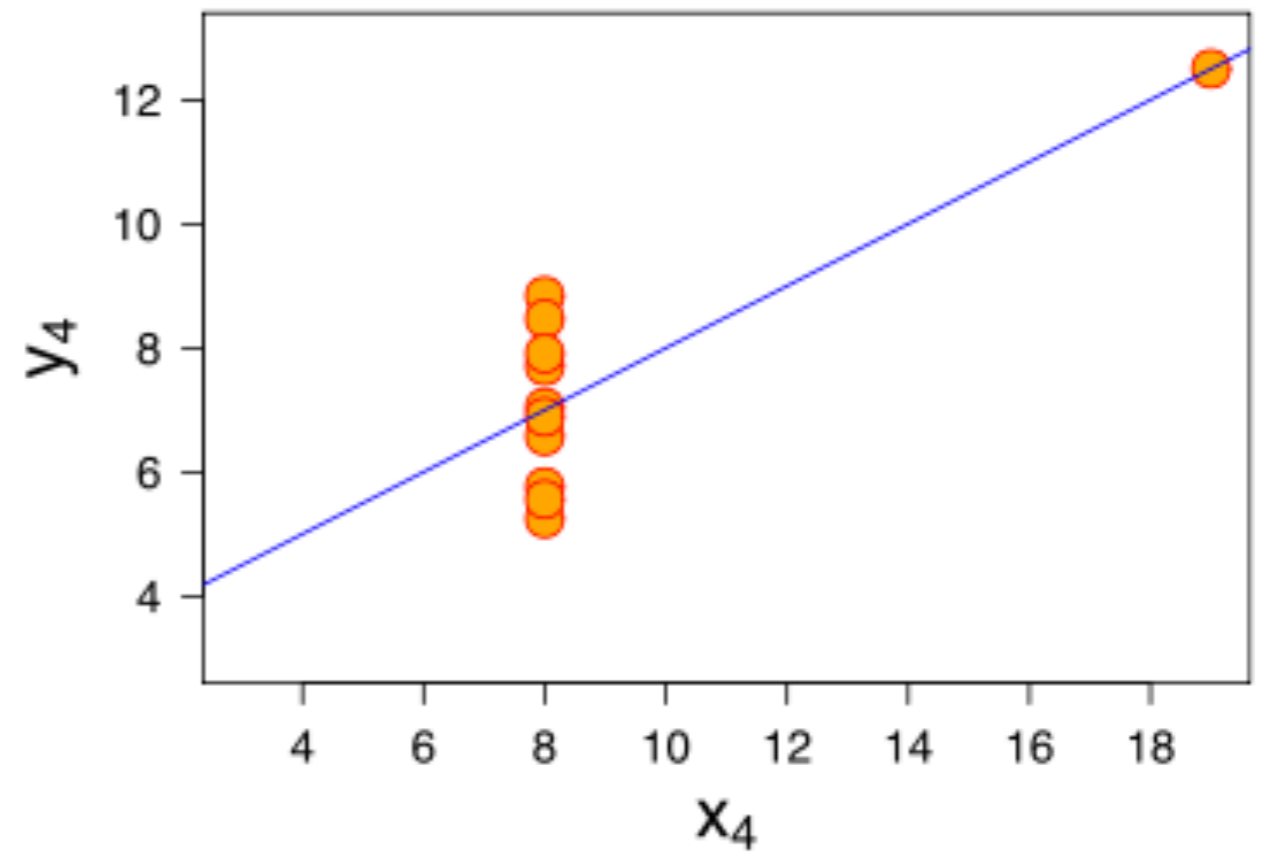
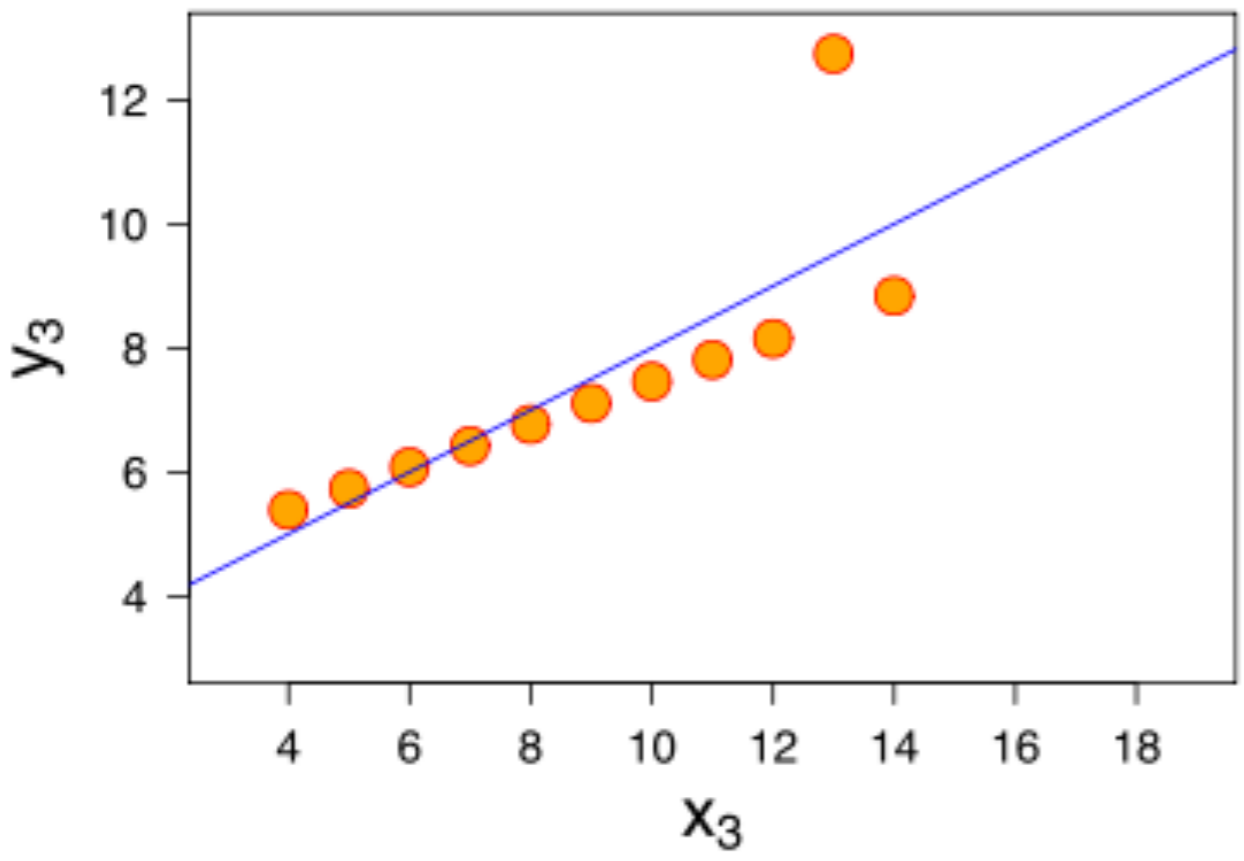
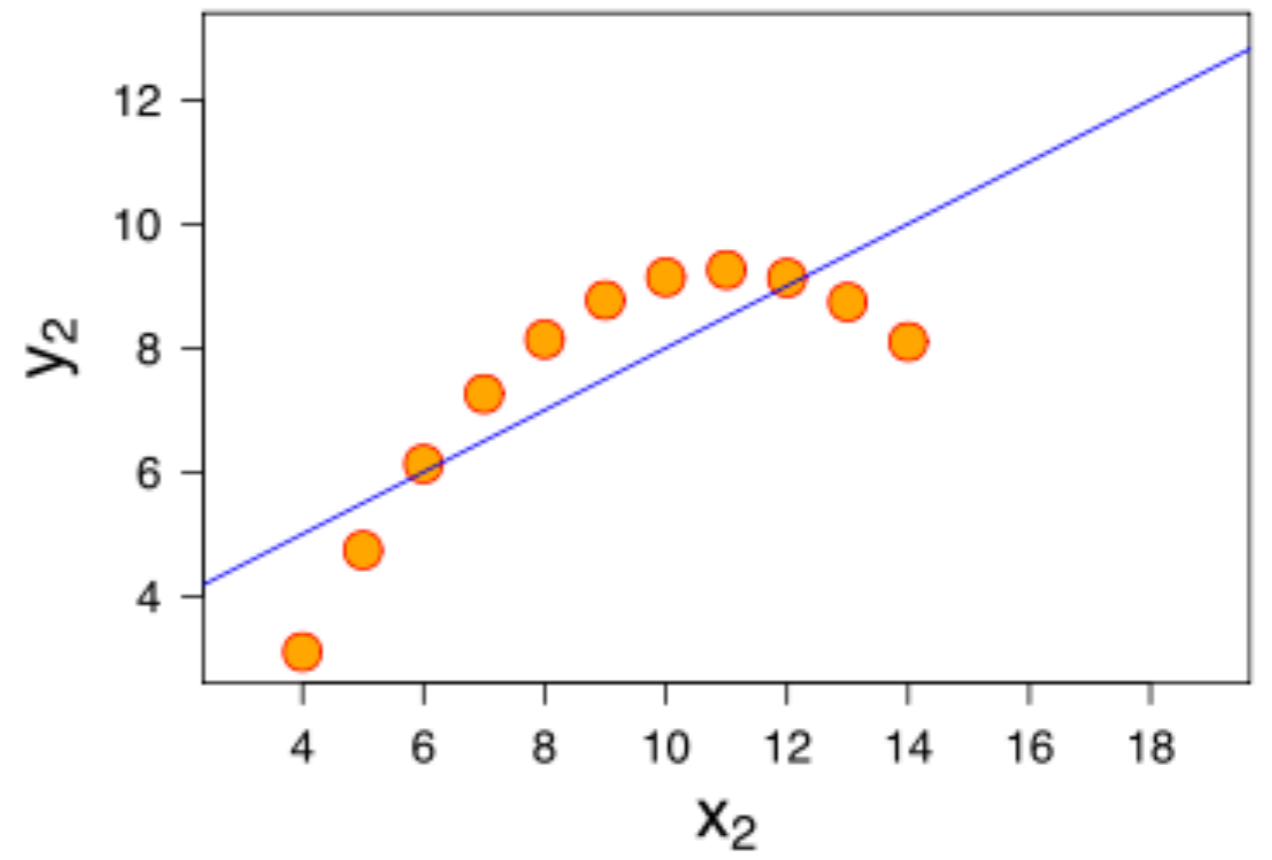
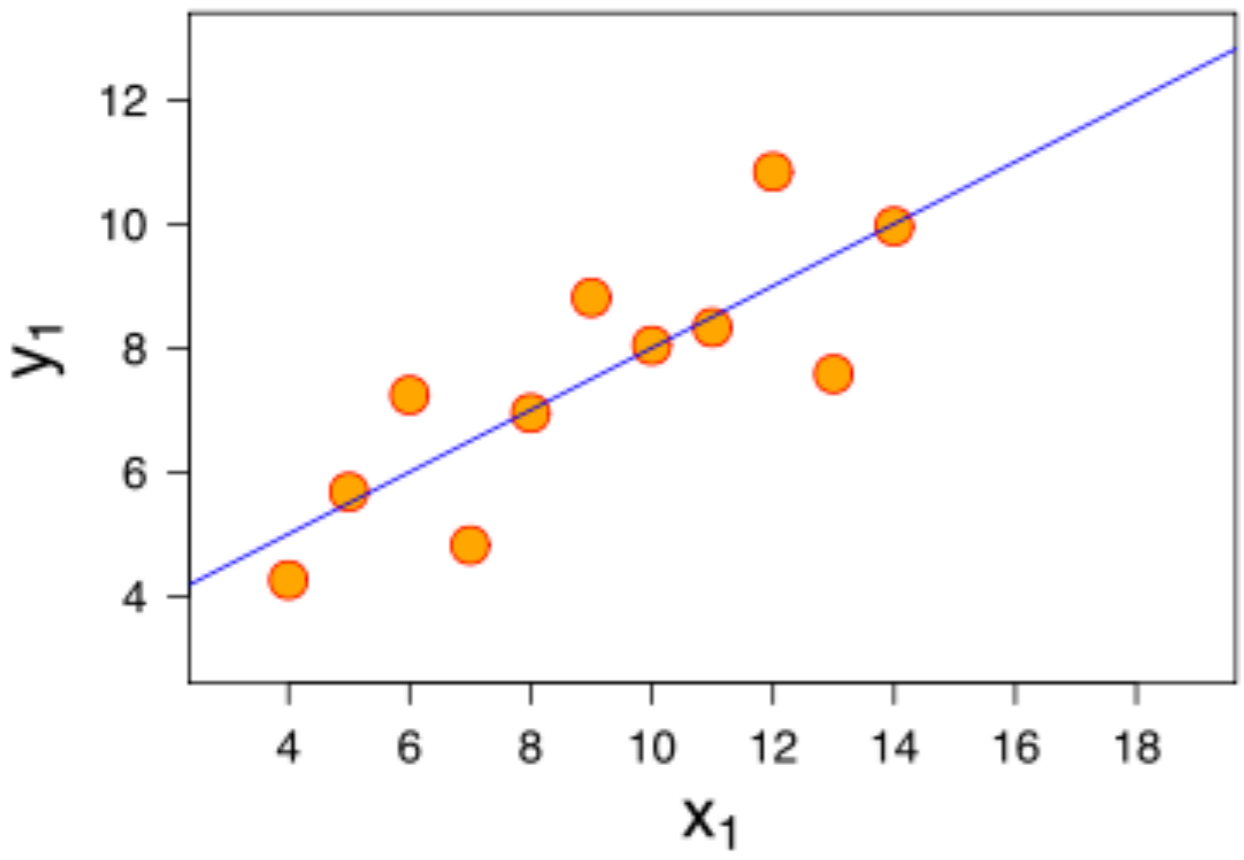
Next

Yield curve 101

The yield curve shows how much it costs the federal government to borrow money for a given amount of time, revealing the relationship between long- and short-term interest rates.

It is, inherently, a forecast for what the economy holds in the future — how much inflation there will be, for example, and how healthy growth will be over the years ahead — all embodied in the price of money today, tomorrow and many years from now.





DEMO

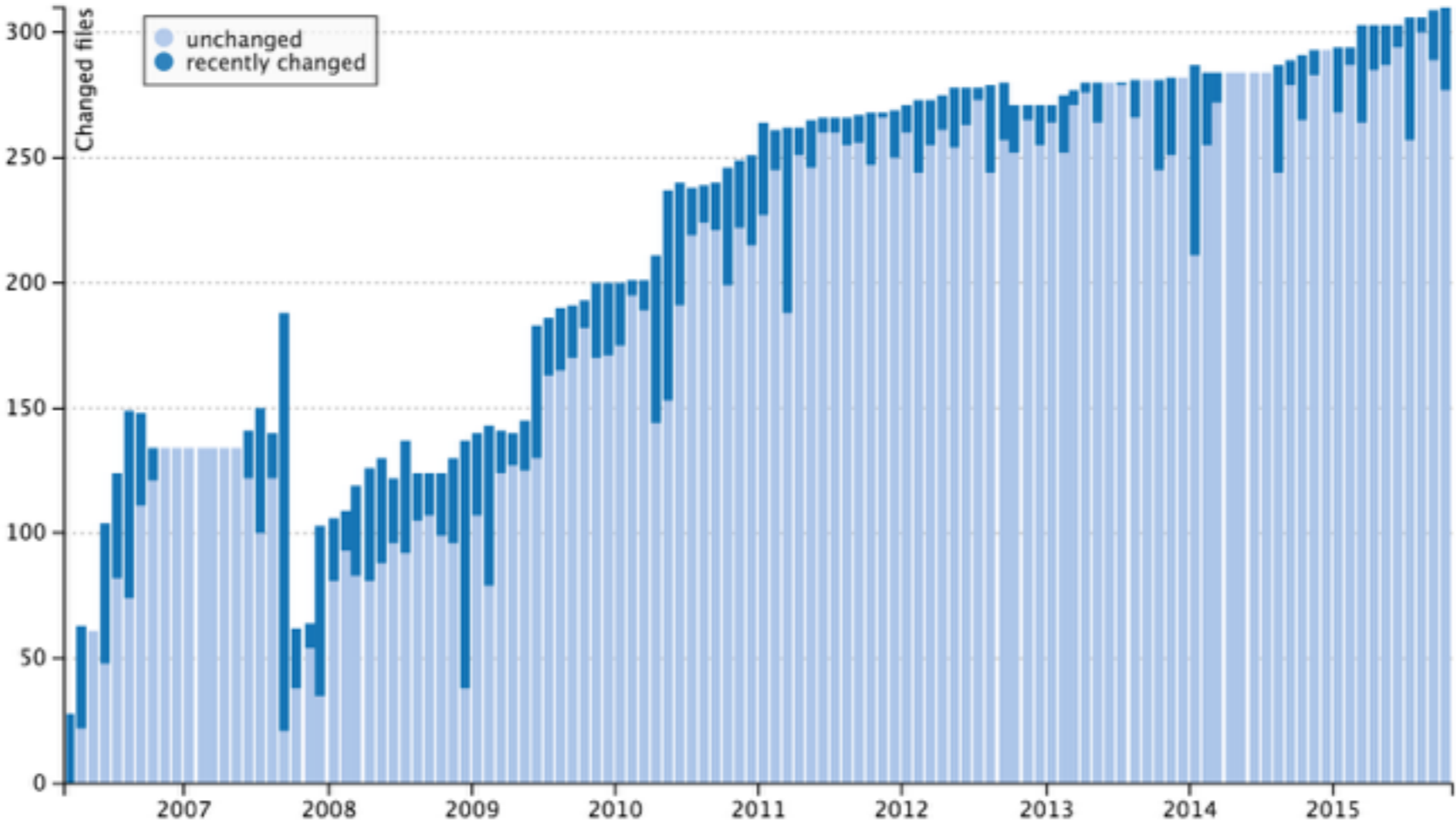
Code History Miner

Server for project code history analysis and visualization.

[Download](#)

0.1 beta, built on 20 Mar 2016

Amount of changing files chart (Clojure)





Plugins

[+ Add new plugin](#)

[Plugins home](#) |
 [IntelliJ IDEA](#) |
 [RubyMine](#) |
 [WebStorm](#) |
 [PhpStorm](#) |
 [PyCharm](#) |
 [AppCode](#) |
 [MPS](#) |
 [Android Studio](#) |
 [DataGrip](#) |
 [CLion](#) |
 [Rider](#) |
  [RSS](#)

Search:

Categories

All plugins (1773)

- [Build \(64\)](#)
- [Code Editing \(97\)](#)
- [Code tools \(196\)](#)
- [Custom Languages \(183\)](#)
- [CVS \(4\)](#)
- [Database \(21\)](#)
- [Editor \(119\)](#)
- [Folder \(4\)](#)
- [Formatting \(27\)](#)
- [Framework integration \(131\)](#)
- [Fun Stuff \(42\)](#)
- [Graphics \(11\)](#)
- [Inspection \(43\)](#)
- [Intention \(8\)](#)
- [J2ee \(20\)](#)
- [J2ME \(3\)](#)
- [JavaScript \(25\)](#)
- [JBehave \(5\)](#)
- [JSP \(3\)](#)
- [Maven 2 \(21\)](#)
- [Menu Components \(7\)](#)

Code History Mining

Products: [IntelliJ IDEA](#), [RubyMine](#), [WebStorm](#), [PhpStorm](#), [PyCharm](#), [AppCode](#), [Android Studio](#), [0xDBE](#), [CLion](#), [Rider](#)

Vendor: [Dmitry Kandalov](#)

Email: [dmitry.kandalov \[at\] gmail.com](mailto:dmitry.kandalov@gmail.com)

Source code: <https://github.com/dkandalov/code-history-mining>

Authors:

[syl](#)

This is proof-of-concept plugin for code history visualization and analysis. For more details and examples of visualizations see [GitHub page](#).

See also [code history miner](#) web server and CLI with similar functionality.

Downloads: **19615**

Rating: 

Participated in rating: **2**

[Comment and rate](#)

Version	Compatible builds	Date	
0.3.0 beta	142.2084 +	03.03.2016	Download
0.2.8 beta	140.2493 — 142.2084	10.08.2015	Download
0.2.6 beta	134 — 140.2493	18.04.2015	Download
0.2.0 beta	122.600 — 133.*	28.03.2014	Download
0.1.7 beta	122.600 +	04.03.2014	Download

18 updates are hidden [show all updates](#)

Project

m junit x

- ▶ doc
- ▶ history
- ▶ lib
- ▶ pmip
- ▶ src
- ▶ target
- hamcrest-all-1.0.jar
- epi-v10.html
- junit.iml
- CONTRIBUTING.md
- README.md
- .classpath
- .gitignore
- .project
- acknowledgements.txt
- BUILDING
- CODING_STYLE.txt
- LEGACY_CODING_STYLE.txt
- LICENSE-junit.txt
- NOTICE.txt
- to-do.txt
- m pom.xml
- .travis.yml

project

```
<?xml version="1.0" encoding="UTF-8"?>
<project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.org/xsd/maven-4.0.0.xsd">
  <modelVersion>4.0.0</modelVersion>

  <groupId>junit</groupId>
  <artifactId>junit</artifactId>
  <version>4.13-SNAPSHOT</version>

  <name>JUnit</name>
  <description>JUnit is a unit testing framework for Java, c</description>
  <url>http://junit.org</url>
  <inceptionYear>2002</inceptionYear>
  <organization>
    <name>JUnit</name>
    <url>http://www.junit.org</url>
  </organization>
  <licenses>
    <license>
      <name>Eclipse Public License 1.0</name>
      <url>http://www.eclipse.org/legal/epl-v10.html</url>
      <distribution>repo</distribution>
    </license>
  </licenses>
</project>
```

Project

- doc
- history
- lib
- pmip
- src
- target
- hamcrest-all-1.0.jar
- epl-v10.html
- junit.iml
- CONTRIBUTING.md
- README.md
- .classpath
- .gitignore
- .project
- acknowledgements.txt
- BUILDING
- CODING_STYLE.txt
- LEGACY_CODING_STYLE.txt
- LICENSE-junit.txt
- NOTICE.txt
- to-do.txt
- m pom.xml**
- .travis.yml

External Libraries

project

```
<?xml version="1.0" encoding="UTF-8"?>  
<project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.org/xsd/maven-4.0.0.xsd">  
  <modelVersion>4.0.0</modelVersion>  
  
  <groupId>junit</groupId>  
  <artifactId>junit</artifactId>  
  <version>4.13-SNAPSHOT</version>  
  
  <name>JUnit</name>  
  <description>JUnit testing framework for Java, C, and Objective-C</description>  
  <url>http://www.junit.org</url>  
  <inceptionYear>2001</inceptionYear>  
  <organization>  
    <name>JUnit</name>  
    <url>http://www.junit.org</url>  
  </organization>  
  <licenses>  
    <license>  
      <name>Eclipse Public License 1.0</name>  
      <url>http://www.eclipse.org/legal/epl-v10.html</url>  
      <distribution>repo</distribution>  
    </license>  
  </licenses>  
</project>
```

Code History Mining

- Grab Project History**
- Current File History Stats
- Amount of Files in Project
- Read Me (page on GitHub)

- ▶ doc
- ▶ history
- ▶ lib
- ▶ pmip
- ▶ src

```
project
<?xml version="1.0" encoding="UTF-8"?>
<project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="
  <modelVersion>4.0.0</modelVersion>
```

Grab History Of Current Project

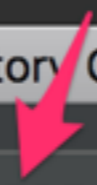
From: To:

Save to:

Grab history on VCS update:

Grab change size in lines/characters and amount of TODOs:

(Please note that grabbing history can slow down IDE and/or take a really long time.)



- LEGACY_CODING_ST
- LICENSE-junit.txt
- NOTICE.txt
- to-do.txt

m pom.xml

.travis.yml

External Libraries

```
<license>
  <name>Eclipse Public License 1.0</name>
  <url>http://www.eclipse.org/legal/epl-v10.html</url>
  <distribution>repo</distribution>
</license>
</licenses>
```

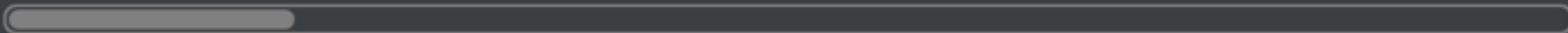
- ▶ doc
- ▶ history
- ▶ lib
- ▶ pmip
- ▶ src
- ▶ target
- ham
- epi-
- junit
- CON
- REAL
- .clas
- .gitic
- .proj
- ackn
- BUILDING
- CODING_STYLE.txt
- LEGACY_CODING_ST
- LICENSE-junit.txt
- NOTICE.txt
- to-do.txt
- m pom.xml
- .travis.yml
- External Libraries


project

```
<?xml version="1.0" encoding="UTF-8"?>  
<project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.org/xsd/maven-4.0.0.xsd">  
  <modelVersion>4.0.0</modelVersion>  
  <groupId>junit</groupId>  
  <artifactId>junit</artifactId>
```

Background Tasks

Grabbing project history





```
</organization>  
<licenses>  
  <license>  
    <name>Eclipse Public License 1.0</name>  
    <url>http://www.eclipse.org/legal/epl-v10.html</url>  
    <distribution>repo</distribution>  
  </license>  
</licenses>
```




Code History Mining

Grabbed history to [junit-file-events.csv](#).
It contains history from 08/07/2001 to 18/04/2016.

You can run [all vizualizations](#) or choose one in plugin popup menu.

projec

<?xml

<proje

<m

<groupId>junit</groupId>

<artifactId>junit</artifactId>

<version>4.13-SNAPSHOT</version>

<name>JUnit</name>

<description>JUnit is a unit testing framework for Java, c

<url>http://junit.org</url>

<inceptionYear>2002</inceptionYear>

<organization>

<name>JUnit</name>

<url>http://www.junit.org</url>

</organization>

<licenses>

<license>

<name>Eclipse Public License 1.0</name>

<url>http://www.eclipse.org/legal/epl-v10.html</url>

<distribution>repo</distribution>

</license>

</licenses>

Project

m junit x

- ▶ doc
- ▶ history
- ▶ lib
- ▶ pmip
- ▶ src
- ▶ target
- hamcrest-all-1.0.jar
- ep1-v10.html
- junit.iml
- CONTRIBUTING.md
- README.md
- .classpath
- .gitignore
- .project
- acknowledgements.txt
- BUILDING
- CODING_STYLE.txt
- LEGACY_CODING_ST
- LICENSE-junit.txt
- NOTICE.txt
- to-do.txt
- m pom.xml
- .travis.yml

External Libraries

Code History ...

Change size: 0/80



00:00

Activity tracker: on

4:5

LF↕

UTF-8

Git: master↕



5

1127 of 1981M

2001-10-21 03:30:18 +0100,8817825612894d11856312a777281b6c7cf81672,kbeck <kbeck>,TestRunner.java,,/juni
2001-10-21 03:30:18 +0100,8817825612894d11856312a777281b6c7cf81672,kbeck <kbeck>,BaseTestRunner.java,,/
2001-10-21 03:30:18 +0100,8817825612894d11856312a777281b6c7cf81672,kbeck <kbeck>,TestRunListener.java,,/
2001-10-21 03:30:18 +0100,8817825612894d11856312a777281b6c7cf81672,kbeck <kbeck>,TestRunner.java,,/juni
2001-10-21 03:30:18 +0100,8817825612894d11856312a777281b6c7cf81672,kbeck <kbeck>,TextRunnerTest.java,,/
2001-10-21 03:30:18 +0100,8817825612894d11856312a777281b6c7cf81672,kbeck <kbeck>,TestRunner.java,,/juni
2001-10-20 01:47:47 +0100,29416b4ee89314119fb0f2d8011d63ac36531388,kbeck <kbeck>,BaseTestRunner.java,,/
2001-10-20 01:47:47 +0100,29416b4ee89314119fb0f2d8011d63ac36531388,kbeck <kbeck>,excluded.properties,,/
2001-10-20 00:48:02 +0100,350d4a4466e9ff618e4a890a424e4a3d6bca324e,kbeck <kbeck>,TextRunnerTest.java,,/
2001-10-19 22:48:32 +0100,80bf23846bfe96f84e7e44436709781262cc1e14,kbeck <kbeck>,Assert.java,,/junit/fr
2001-10-19 22:48:32 +0100,80bf23846bfe96f84e7e44436709781262cc1e14,kbeck <kbeck>,ClassPathTestCollector
2001-10-19 22:48:32 +0100,80bf23846bfe96f84e7e44436709781262cc1e14,kbeck <kbeck>,AllTests.java,,/junit/
2001-10-19 22:48:32 +0100,80bf23846bfe96f84e7e44436709781262cc1e14,kbeck <kbeck>,AssertTest.java,,/juni
2001-10-19 22:48:32 +0100,80bf23846bfe96f84e7e44436709781262cc1e14,kbeck <kbeck>,SimpleTestCollectorTes
2001-10-02 20:38:22 +0100,0bb3dfe2939cc214ee5e77556a48d4aea9c6396a,kbeck <kbeck>,IMoney.java,,/junit/sa
2001-10-02 20:38:22 +0100,0bb3dfe2939cc214ee5e77556a48d4aea9c6396a,kbeck <kbeck>,Money.java,,/junit/sam
2001-10-02 20:38:22 +0100,0bb3dfe2939cc214ee5e77556a48d4aea9c6396a,kbeck <kbeck>,MoneyBag.java,,/junit/
2001-10-02 20:38:22 +0100,0bb3dfe2939cc214ee5e77556a48d4aea9c6396a,kbeck <kbeck>,MoneyTest.java,,/junit
2001-07-09 23:51:53 +0100,ce0bb8f59ea7de1ac3bb4f678f7ddf84fe9388ed,egamma <egamma>,,.classpath,,NEW,-1,
2001-07-09 23:51:53 +0100,ce0bb8f59ea7de1ac3bb4f678f7ddf84fe9388ed,egamma <egamma>,,.vcm_meta,,,MODIFICA
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,IMG00001.GIF,,/doc/c
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,IMG00002.GIF,,/doc/c
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,logo.gif,,/doc/cookb
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,Image1.gif,,/doc/coo
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,Image2.gif,,/doc/coo
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,Image3.gif,,/doc/coo
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,Image4.gif,,/doc/coo
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,Image5.gif,,/doc/coo
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,Image6.gif,,/doc/coo
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,Image7.gif,,/doc/coo
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,IMG00001.GIF,,/doc/t
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,IMG00002.GIF,,/doc/t
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,IMG00003.GIF,,/doc/t
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,logo.gif,,/doc/testi
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,logo.gif,,/junit/run
2001-07-09 23:16:13 +0100,4b4c3992e4547b4d00ea490f03398f2d0e5c8f7e,egamma <egamma>,,.vcm_meta,,,NEW,-1,-
2001-07-08 20:42:20 +0100,2fd22847241b16be7d79a1f9de02957168aa13b9,egamma <egamma>,test.jar,,/junit/test
2001-05-21 21:09:24 +0100,8076b782f8c9ac1df6569bbfa479db2da9e6cbcf,egamma <egamma>,Version.java,,/junit

```
2001-10-21 03:30:18 +0100,8817825612894d11856312a777281b6c7cf81672,kbeck <kbeck>,TestRunner.java,,/juni
2001-10-21 03:30:18 +0100,8817825612894d11856312a777281b6c7cf81672,kbeck <kbeck>,BaseTestRunner.java,,/
2001-10-21 03:30:18 +0100,8817825612894d11856312a777281b6c7cf81672,kbeck <kbeck>,TestRunListener.java,,
2001-10-21 03:30:18 +0100,8817825612894d11856312a777281b6c7cf81672,kbeck <kbeck>,TestRunner.java,,/juni
2001-10-21 03:30:18 +0100,8817825612894d11856312a777281b6c7cf81672,kbeck <kbeck>,TextRunnerTest.java,,/
2001-10-21 03:30:18 +0100,8817825612894d11856312a777281b6c7cf81672,kbeck <kbeck>,TestRunner.java,,/juni
2001-10-20 01:47:47 +0100,29416b4ee89314119fb0f2d8011d63ac36531388,kbeck <kbeck>,BaseTestRunner.java,,/
2001-10-20 01:47:47 +0100,29416b4ee89314119fb0f2d8011d63ac36531388,kbeck <kbeck>,excluded.properties,,/
2001-10-20 00:48:02 +0100,80bf23846bfe96f84e7e44436709781262cc1e14,kbeck <kbeck>,TextRunnerTest.java,,/
2001-10-19 22:48:32 +0100,80bf23846bfe96f84e7e44436709781262cc1e14,kbeck <kbeck>,Assert.java,,/junit/fr
2001-10-19 22:48:32 +0100,80bf23846bfe96f84e7e44436709781262cc1e14,kbeck <kbeck>,ClassPathTestCollector
2001-10-19 22:48:32 +0100,80bf23846bfe96f84e7e44436709781262cc1e14,kbeck <kbeck>,AllTests.java,,/junit/
2001-10-19 22:48:32 +0100,80bf23846bfe96f84e7e44436709781262cc1e14,kbeck <kbeck>,AssertTest.java,,/juni
2001-10-19 22:48:32 +0100,80bf23846bfe96f84e7e44436709781262cc1e14,kbeck <kbeck>,SimpleTestCollectorTes
2001-10-02 20:38:22 +0100,0bb3dfe2939cc214ee5e77556a48d4aea9c6396a,kbeck <kbeck>,IMoney.java,,/junit/sa
2001-10-02 20:38:22 +0100,0bb3dfe2939cc214ee5e77556a48d4aea9c6396a,kbeck <kbeck>,Money.java,,/junit/sam
2001-10-02 20:38:22 +0100,0bb3dfe2939cc214ee5e77556a48d4aea9c6396a,kbeck <kbeck>,MoneyBag.java,,/junit/
2001-10-02 20:38:22 +0100,0bb3dfe2939cc214ee5e77556a48d4aea9c6396a,kbeck <kbeck>,MoneyTest.java,,/junit
2001-07-09 23:51:53 +0100,ce0bb8f59ea7de1ac3bb4f678f7ddf84fe9388ed,egamma <egamma>,,.classpath,,NEW,-1,
2001-07-09 23:51:53 +0100,ce0bb8f59ea7de1ac3bb4f678f7ddf84fe9388ed,egamma <egamma>,,.vcm_meta,,,MODIFICA
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,IMG00001.GIF,,/doc/c
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,IMG00002.GIF,,/doc/c
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,logo.gif,,/doc/cookb
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,Image1.gif,,/doc/coo
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,Image2.gif,,/doc/coo
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,Image3.gif,,/doc/coo
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,Image4.gif,,/doc/coo
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,Image5.gif,,/doc/coo
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,Image6.gif,,/doc/coo
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,Image7.gif,,/doc/coo
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,IMG00001.GIF,,/doc/t
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,IMG00002.GIF,,/doc/t
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,IMG00003.GIF,,/doc/t
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,logo.gif,,/doc/testi
2001-07-09 23:39:54 +0100,3216b428b87523b6b9247bd9d84c36c2ded3a3fa,egamma <egamma>,logo.gif,,/junit/run
2001-07-09 23:16:13 +0100,4b4c3992e4547b4d00ea490f03398f2d0e5c8f7e,egamma <egamma>,,.vcm_meta,,,NEW,-1,-
2001-07-08 20:42:20 +0100,2fd22847241b16be7d79a1f9de02957168aa13b9,egamma <egamma>,test.jar,,/junit/test
2001-05-21 21:09:24 +0100,8076b782f8c9ac1df6569bbfa479db2da9e6cbcf,egamma <egamma>,Version.java,,/junit
```

Project

m junit x

- ▶ doc
- ▶ history
- ▶ lib
- ▶ pmip
- ▶ src
- ▶ target
- hamcrest-all-1.0.jar
- epi-v10.html
- junit.iml
- CONTRIBUTING.md
- README.md
- .classpath
- .gitignore
- .project
- acknowledgements.txt
- BUILDING
- CODING_STYLE.txt
- LEGACY_CODING_STYLE.txt
- LICENSE-junit.txt
- NOTICE.txt
- to-do.txt
- m pom.xml
- .travis.yml

project

```
<?xml version="1.0" encoding="UTF-8"?>
<project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.org/xsd/maven-4.0.0.xsd">
  <modelVersion>4.0.0</modelVersion>

  <groupId>junit</groupId>
  <artifactId>junit</artifactId>
  <version>4.13-SNAPSHOT</version>

  <name>JUnit</name>
  <description>JUnit is a unit testing framework for Java, c</description>
  <url>http://junit.org</url>
  <inceptionYear>2002</inceptionYear>
  <organization>
    <name>JUnit</name>
    <url>http://www.junit.org</url>
  </organization>
  <licenses>
    <license>
      <name>Eclipse Public License 1.0</name>
      <url>http://www.eclipse.org/legal/epl-v10.html</url>
      <distribution>repo</distribution>
    </license>
  </licenses>
</project>
```

Project view showing file structure:

- doc
- history
- lib
- pmip
- src
- target
- hamcrest-all-1.0.jar
- epl-v10.html
- junit.iml
- CONTRIBUTING.md
- README.md
- .classpath
- .gitignore
- .project
- acknowledgements.txt
- BUILDING
- CODING_STYLE.txt
- LEGACY_CODING_STYLE.txt
- LICENSE-junit.txt
- NOTICE.txt
- to-do.txt
- m pom.xml**
- .travis.yml

External Libraries

project

```
<?xml version="1.0" encoding="UTF-8"?>
<project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  <modelVersion>4.0.0</modelVersion>

  <groupId>junit</groupId>
  <artifactId>junit</artifactId>
  <version>4.13-SNAPSHOT</version>
  <name>junit</name>
  <description>junit testing framework for Java, C, and Objective-C</description>
  <url>http://www.junit.org</url>
  <inceptionYear>2001</inceptionYear>
  <organization>
    <name>JUnit</name>
    <url>http://www.junit.org</url>
  </organization>
  <licenses>
    <license>
      <name>Eclipse Public License 1.0</name>
      <url>http://www.eclipse.org/legal/epl-v10.html</url>
      <distribution>repo</distribution>
    </license>
  </licenses>
</project>
```

Code History Mining

- Grab Project History**
- junit-file-events.csv
- Current File History Stats
- Amount of Files in Project
- Read Me (page on GitHub)

Project

- doc
- history
- lib
- pmip
- src
- target
- hamcrest-all-1.0.jar
- epl-v10.html
- junit.iml
- CONTRIBUTING.md
- README.md
- .classpath
- .gitignore
- .project
- acknowledgements.txt
- BUILDING
- CODING_STYLE.txt
- LEGACY_CODING_STYLE.txt
- LICENSE-junit.txt
- NOTICE.txt
- to-do.txt
- m pom.xml**
- .travis.yml

External Libraries

project

```

<?xml version="1.0" encoding="UTF-8"?>
<project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.org/xsd/maven-4.0.0.xsd">
  <modelVersion>4.0.0</modelVersion>
  <groupId>junit</groupId>
  <artifactId>junit</artifactId>
  <version>4.13-SNAPSHOT</version>
  <name>JUnit</name>
  <description>JUnit is a Java class that can be used to write and run tests. It is a member of the Eclipse Public License (EPL) and is distributed under the terms of the EPL. It is a member of the Eclipse Public License (EPL) and is distributed under the terms of the EPL.
  <url>http://www.junit.org</url>
  <organization>
    <name>JUnit</name>
    <url>http://www.junit.org</url>
  </organization>
  <licenses>
    <license>
      <name>Eclipse Public License</name>
      <url>http://www.eclipse.org/legal/epl-v10.html</url>
      <distribution>repo</distribution>
    </license>
  </licenses>
</project>

```

- Code History Mining
- Grab Project History
- junit-file-events.csv**
 - All Visualizations
 - Open Script Editor**
 - Code Churn Chart
 - Amount Of Committers Chart
 - Amount Of Commits By Committer
 - Amount Of TODOs Chart
 - Amount Of Files In Commit Chart
 - Amount Of Changing Files Chart
 - Change Size By File Type Chart
 - Changes Treemap
 - Files In The Same Commit Graph
 - Committers Changing Same Files Graph
 - Commit Time Punchcard
 - Commit Messages Word Chart
 - Commit Messages Word Cloud
- Current File History Stats
- Amount of Files in Project
- Read Me (page on GitHub)

- ▼ **junit** ~/IdeaProjects/ju
- ▶ .idea
- ▶ .settings
- ▶ doc
- ▶ history
- ▶ lib
- ▶ pmip
- ▶ src
- ▶ target
 - hamcrest-all-1.0.jar
 - epl-v10.html
 - junit.iml
 - CONTRIBUTING.md
 - README.md
 - .classpath
 - .gitignore
 - .project
 - acknowledgements.txt
 - BUILDING
 - CODING_STYLE.txt
 - LEGACY_CODING_ST
 - LICENSE-junit.txt
 - NOTICE.txt
 - to-do.txt

```
// To run the script use alt+shift+E (or "Run Code History Sc  
// For more details about scripts and examples see GitHub wiki  
// https://github.com/dkandalov/code-history-mining/wiki/Code-  
  
data.size()
```

- ▼ junit ~/IdeaProjects/ju
 - ▶ .idea
 - ▶ .settings
 - ▶ doc
 - ▶ history
 - ▶ lib
 - ▶ pmip
 - ▶ src
 - ▶ target
 - hamcrest-all-1.0.jar
 - epl-v10.html
 - junit.iml
 - CONTRIBUTING.md
 - README.md
 - .classpath
 - .gitignore

```
// To run the script use alt+shift+E (or "Run Code History Sc
// For more details about scripts and examples see GitHub wiki
// https://github.com/dkandalov/code-history-mining/wiki/Code-
data.size()
```

Event Log

13:33:55 7306

Event Log content area with various icons and a right arrow button.


```
// To run the script use alt+shift+E (or "Run Code History Script" in  
// For more details about scripts and examples see GitHub wiki  
// https://github.com/dkandalov/code-history-mining/wiki/Code-History
```

```
data.take(10)
```

```
commitTime,revision,author,fileNameBefore,fileName,packageNameBefore,packageName,fileChangeType,commitMessage
2016-04-18 21:45:19 +0100,fd691746a7d953c8e4a6855161437124071a49fb,Pascal Schumacher,,ReleaseNotes4.5.html,,/
2016-04-18 21:45:19 +0100,fd691746a7d953c8e4a6855161437124071a49fb,Pascal Schumacher,,ReleaseNotes4.4.md,,/do
2016-04-18 21:45:19 +0100,fd691746a7d953c8e4a6855161437124071a49fb,Pascal Schumacher,,ReleaseNotes4.4.html,,/
2016-04-18 15:15:43 +0100,0431f42c2f2502610719e0b45a28fd0dd51b2cb1,Alberto Scotto,,ParentRunner.java,,/src/ma
2016-04-05 18:51:15 +0100,82bea5dc259bb769fcd590020fe42733c95e3cfa,Paul Duffin,,TestWatchmanTest.java,,/src/t
2016-04-05 18:51:15 +0100,82bea5dc259bb769fcd590020fe42733c95e3cfa,Paul Duffin,,SimpleTest.java,,/src/test/ja
2016-04-05 18:51:15 +0100,82bea5dc259bb769fcd590020fe42733c95e3cfa,Paul Duffin,,AllValidatorTests.java,,/src/
2016-04-05 18:51:15 +0100,82bea5dc259bb769fcd590020fe42733c95e3cfa,Paul Duffin,,AllTheoriesTests.java,,/src/t
2016-04-05 18:51:15 +0100,82bea5dc259bb769fcd590020fe42733c95e3cfa,Paul Duffin,,AllSamplesTests.java,,/src/te
2016-04-05 18:51:15 +0100,82bea5dc259bb769fcd590020fe42733c95e3cfa,Paul Duffin,,AllRulesTests.java,,/src/test
```

```
// To run the script use alt+shift+E (or "Run Code History Script" in  
// For more details about scripts and examples see GitHub wiki  
// https://github.com/dkandalov/code-history-mining/wiki/Code-History
```

```
data.collect{ it.author }.unique()
```

```
// To run the script use alt+shift+E (or "Run Code History Script" in  
// For more details about scripts and examples see GitHub wiki  
// https://github.com/dkandalov/code-history-mining/wiki/Code-History
```

```
data.collect{ it.author }.unique()
```

Event Log

```
13:45:10 Kristian Rosenvold  
13:45:10 KentBeck  
13:45:10 alanharder  
13:45:10 watisdit  
13:45:10 Jon Kean  
13:45:10 James Abley  
13:45:10 Mike Forsberg  
13:45:10 kbeck  
13:45:10 cmschmidt  
13:45:10 clarkware  
13:45:10 egamma  
13:45:10 emeade  
13:45:10 vbossica|
```

DATA CLEANSING

“NON-ESSENTIAL” FILES

pom.xml

CHANGELOG

generated files

library source code

various binary artefacts

SUPERFICIAL CHANGES

VERSION CONTROL BOTS

**CHANGED COMMITTER NAME
AND PAIR PROGRAMMING**

OTHER DATA ISSUES

file move as delete/add

git rebase

mistakes in commits

corrupted repositories

HOW TO CLEAN?

check data manually

run analysis and check for sanity

visualize

DIY

READ FROM VERSION CONTROL

<https://github.com/dkandalov/vcs-reader>

<https://github.com/codeinn/vcs>

other VCS-specific libraries

parse CLI stdout

```
// setup project
def gitSettings = GitSettings.defaults()
def vcsRoot = new GitVcsRoot(
    "/tmp/junit",
    "https://github.com/junit-team/junit",
    gitSettings
)
def vcsProject = new VcsProject(vcsRoot)
```

```
// clone from GitHub
def cloneResult = vcsProject.cloneToLocal()
assert cloneResult.isSuccessful()

// log commits within date range
def logResult = vcsProject.log(
    date("01/01/2013"),
    date("01/01/2014")
)
assert logResult.commits().size() == 242
```

```
// read commit data
def commit = logResult.commits().last()
assert commit.revision == "0e1a559e1371aa9929ca4f61f87cf8f9a5923ce7"
assert commit.revisionBefore == "4e9f1a65ca8d794db54260b4f2e5b078d949fdda"
assert commit.time == dateTime("Fri Dec 27 18:54:49 GMT 2013")
assert commit.author == "Stefan Birkner"
assert commit.message.contains("Override toString()")
```



```
// read commit changes
def firstChange = commit.changes.first()
assert firstChange.type == MODIFIED
assert firstChange.filePath ==
    "src/main/java/org/junit/runners/model/FrameworkField.java"
assert firstChange.fileContent().value.contains("public class FrameworkField")
```

ANALYSE

general purpose languages

NumPy, pandas, R, etc.

neo4j?

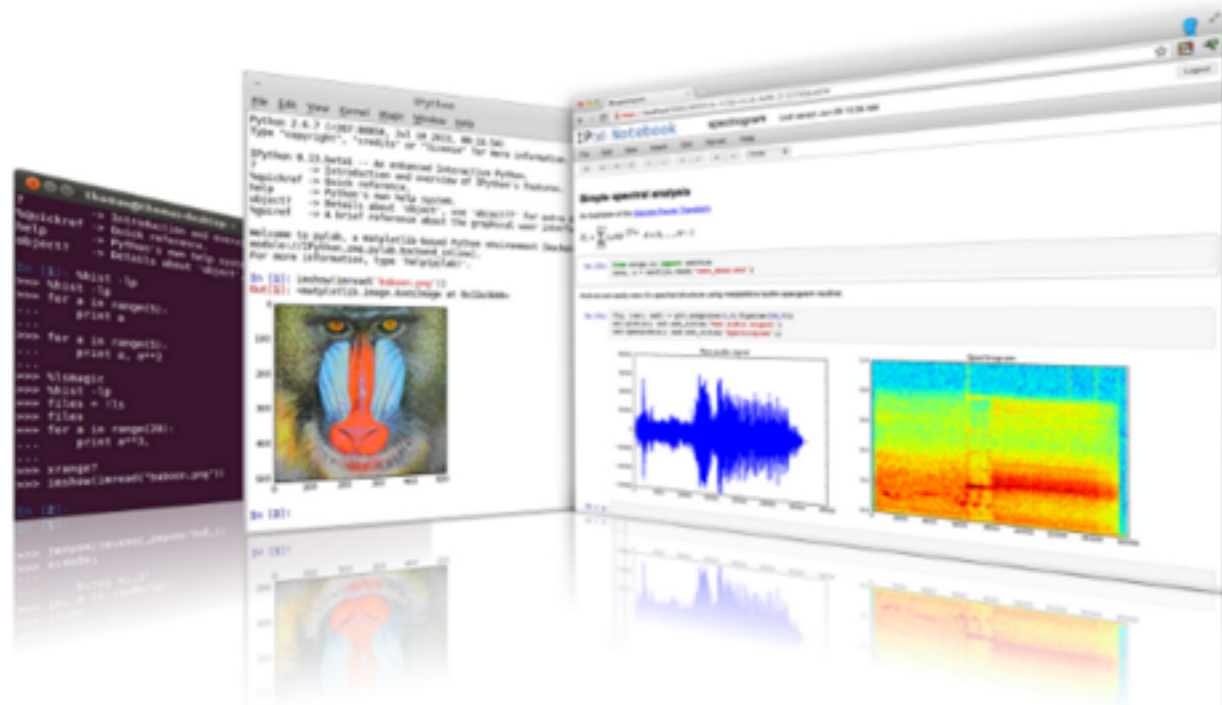
```
def topCommitters = events
  .groupBy({ it.author }, { it.revision })
  .rollup(0){ it.size() }
  .findAll{ it.value > 100 }
  .sort{ -it.value }
  .take(10)*.key
```

IP[y]: IPython Interactive Computing

[Install](#) · [Documentation](#) · [Project](#) · [Jupyter](#) · [News](#) · [Cite](#) · [Donate](#)

IPython provides a rich architecture for interactive computing with:

- A powerful interactive shell.
- A kernel for [Jupyter](#).
- Support for interactive data visualization and use of [GUI toolkits](#).
- Flexible, [embeddable](#) interpreters to load into your own projects.
- Easy to use, high performance tools for [parallel computing](#).



To get started with IPython in the Jupyter Notebook, see our [official example collection](#). Our [notebook gallery](#) is an excellent way to see the many things you can do with IPython while learning about a variety of topics, from basic programming to advanced statistics or quantum mechanics.

```

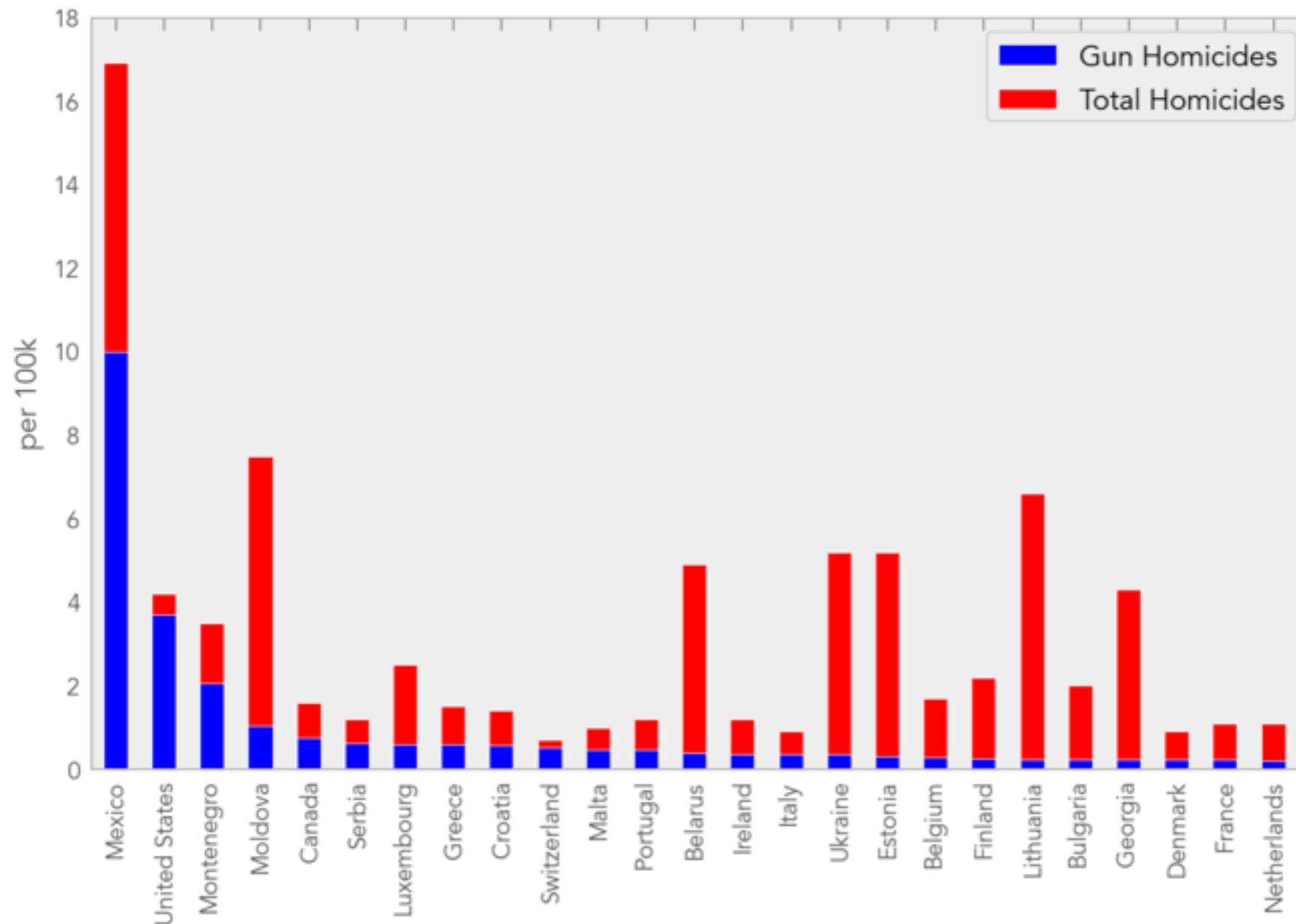
In [13]: index = (data['Region'] == 'Europe') + \
            (data.index == 'United States') + \
            (data.index == 'Canada') + \
            (data.index == 'Mexico')
selected = data[index]

print "By Total Gun Homicides"
sys.stdout.flush()

by_guns = selected.sort("Gun Homicides", ascending=False)
#by_guns['Gun Homicides'].plot(kind='bar')
plot_percapita(by_guns, limit=25)
display_relevant(selected, limit=None)

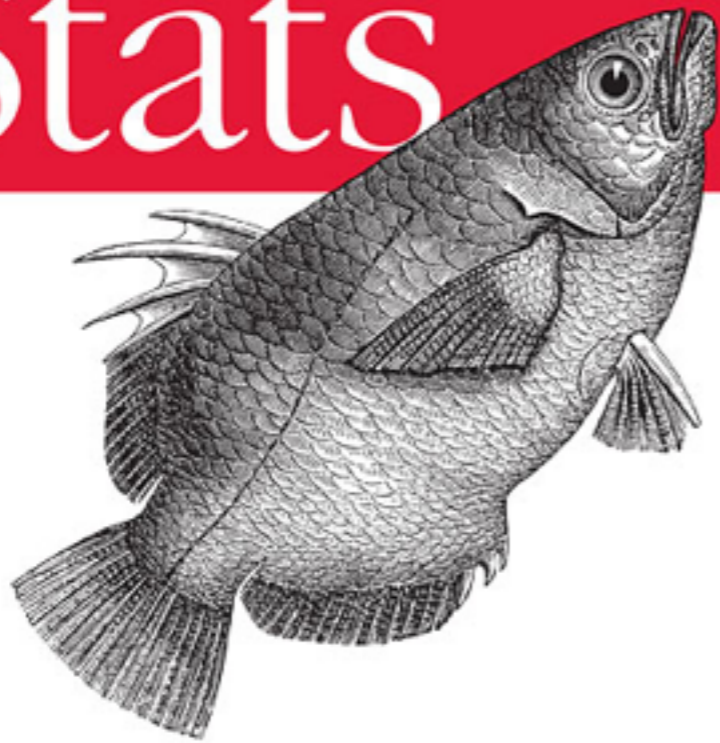
```

By Total Gun Homicides



Probability and Statistics for Programmers

Think Stats



O'REILLY®

Allen B. Downey

Bayesian Statistics in Python

Think Bayes



O'REILLY®

Allen B. Downey

VISUALISE

author,commits,changedFiles
David Saff,314,1225
dsaff,257,2226
kbeck,121,920
Stefan Birkner,109,288
Tibor Digana,100,169
Marc Philipp,86,233
Kevin Cooney,53,154
egamma,48,295
emeade,37,54
pimterry,36,113
Stephen Connolly,24,42
Arie van Deursen,23,24
Noel Yap,23,125
flpa,21,47
clarkware,21,21
Ricky,18,76
Matthew Farwell,18,117
CloudBees DEV@Cloud,14,14
Corey Vaillancourt,14,17
Alistair A. Israel,14,26
Arjan Wulder,13,326
Jesse Glick,13,21
Petter Måhlén,11,17
Adam Beneschan,11,32
Dave Grant,9,9
Brett Chabot,9,13
Juan Cortez,9,16
Andreas Skoog,8,17


```
fileType,changeCount
java,6027
xml,268
txt,232
html,214
class,82
md,67
htm,67
gif,47
prefs,39
jar,36
"",30
ant,26
sh,25
classpath,24
gitignore,17
GIF,16
cvsignore,15
css,11
properties,9
project,9
js,5
template,5
fml,4
yaml,4
zip,4
rb,3
```

```
fileName, fileChanges
"/src/test/java/org/junit/tests/AllTests.java",131
"/build.xml",120
"/src/main/java/org/junit/runners/ParentRunner.java",112
"/src/main/java/org/junit/runners/BlockJUnit4ClassRunner.java",101
"/src/main/java/junit/runner/Version.java",95
"/src/main/java/org/junit/runners/Parameterized.java",93
"/pom.xml",91
"/acknowledgements.txt",74
"/src/main/java/org/junit/experimental/theories/Theories.java",72
"/src/main/java/org/junit/runners/Suite.java",69
"/src/main/java/org/junit/runners/model/TestClass.java",69
"/src/main/java/org/junit/Assert.java",68
"/src/main/java/org/junit/rules/ExpectedException.java",62
"/src/main/java/org/junit/runner/Description.java",55
"/src/test/java/org/junit/tests/assertion/AssertionTest.java",54
"/src/main/java/org/junit/internal/runners/JUnit4ClassRunner.java",46
"/src/main/java/org/junit/experimental/max/MaxCore.java",46
"/doc/ReleaseNotes4.7.txt",46
"/src/main/java/org/junit/runners/model/FrameworkMethod.java",45
"/src/main/java/org/junit/runner/JUnitCore.java",44
"/src/main/java/org/junit/internal/runners/statements/FailOnTimeout.java",42
"/src/main/java/org/junit/runner/Request.java",40
"/src/test/java/org/junit/tests/experimental/max/MaxStarterTest.java",39
"/src/main/java/org/junit/experimental/categories/Categories.java",37
```

fileChanges, revision, author, message
378,478f7f9217cf5da219a7ba298e3a87f405378194,dsaff,"Merged with branch, Kent will
305,30f2b16525dabb477373be9ed3e76bb98b200806,Arjan Wulder,Issue #426 Applied new
249,8226e3e64a7d4ace29eb08ab6c9fc9d4055f20c9,dsaff,"- Moved src directory into tw
221,168fe0844e104a224a5899b02dfd3e0fb07a32e3,kbeck,Final changes for release
171,40cc52cf5f5dc1dd1ee513410536bdc11422307d,dsaff,"Created a separate src folder
71,6acc164b4125da775cccd4035c1b99e9d404bacd,dsaff,"- The hamcrest-core-1.1 library
59,0e1aca2c28858ac1ad77ef40f704ec1b610c9d65,Matthew Farwell,Fixes #67 no since ja
58,df00d5eced3a7737b88de0f6f9e3673f0cf88f88,Alex Yursha,"remove 'f' prefix from f
57,3b5085dec3f1fd578c5109114bd02370bea48fda,kbeck,Fixed two warnings
55,65a98b47a95ff6173c5190d0afff07c7265fe277,Paul Duffin,Clean up the test hierarc
54,35f5f3a78cdc49954e9542dafeea86f13705fdf8,dsaff,Reorganize tests
50,4b1869ebb8002e5d0b82ab55460f6126043c9ec4,egamma,JUnit 3.5 candidate
49,86b71cb1b9784a9b87c86b4f3f12632ae6773d1f,egamma,finished support for no-arg co
47,d322fe6d909c929be5e637b2adbd7da8f9264b3a,dsaff,Reorganization of Theory design
38,b4e915892bef6d4885cbde7db91476a0a9138c4f,cmschmidt,Javadoc improvements: - cod
37,a9815c0f1419e20b6282c3944195083bb38deaa9,kbeck,"4.5 release candidate minus ja
37,ee235bc8c5cf5d4345a690f1bd7e3fd1967ef8ea,dsaff,Re-organize theory packages
36,576377b9ad44028dc84c2f8487cfe8dbccb8c252,dsaff,"- Split links into their own p
34,4e2455b1f142bdf8f07383082638718695e51224,David Saff,Rollback to hamcrest 1.1
34,e35fb32e07a4cfe69dd86024a79d9cb3b9132e55,David Saff,Upgraded to Hamcrest 1.2
34,52e61a55e8ffcaab2586ea8fec28e4b01336cd30,David Saff,Upgraded to Hamcrest 1.2\n
33,d9c908b9aab5f610e2f42aba1863ce25e36423f2,dsaff,API clean-up before 4.5 release
32,d9cccce5034bb9876221d271914471b205220061,Marc Philipp,"Upgrade to Hamcrest 1.3

LIBRARIES

<https://d3js.org>

<http://nvd3.org>

<https://github.com/dc-js/dc.js>

<https://github.com/dkandalov/d3-components>

<http://stanford.edu/~mwaskom/software/seaborn/index.html>



OMG WTF

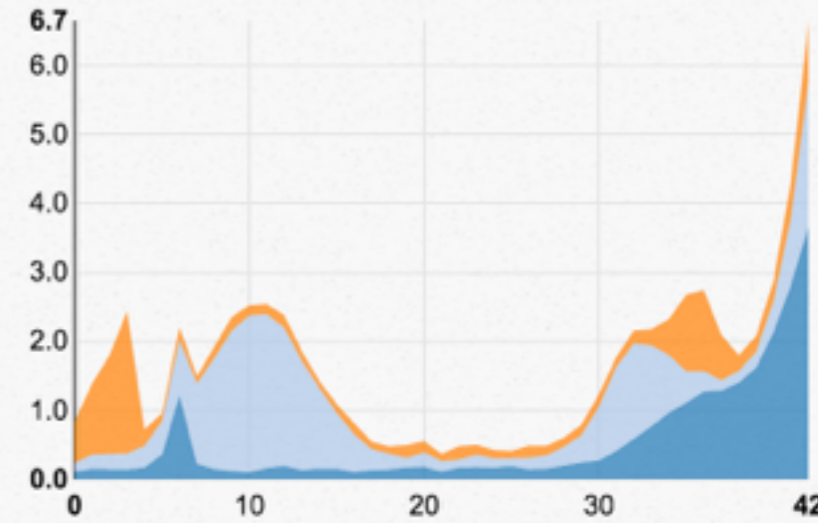
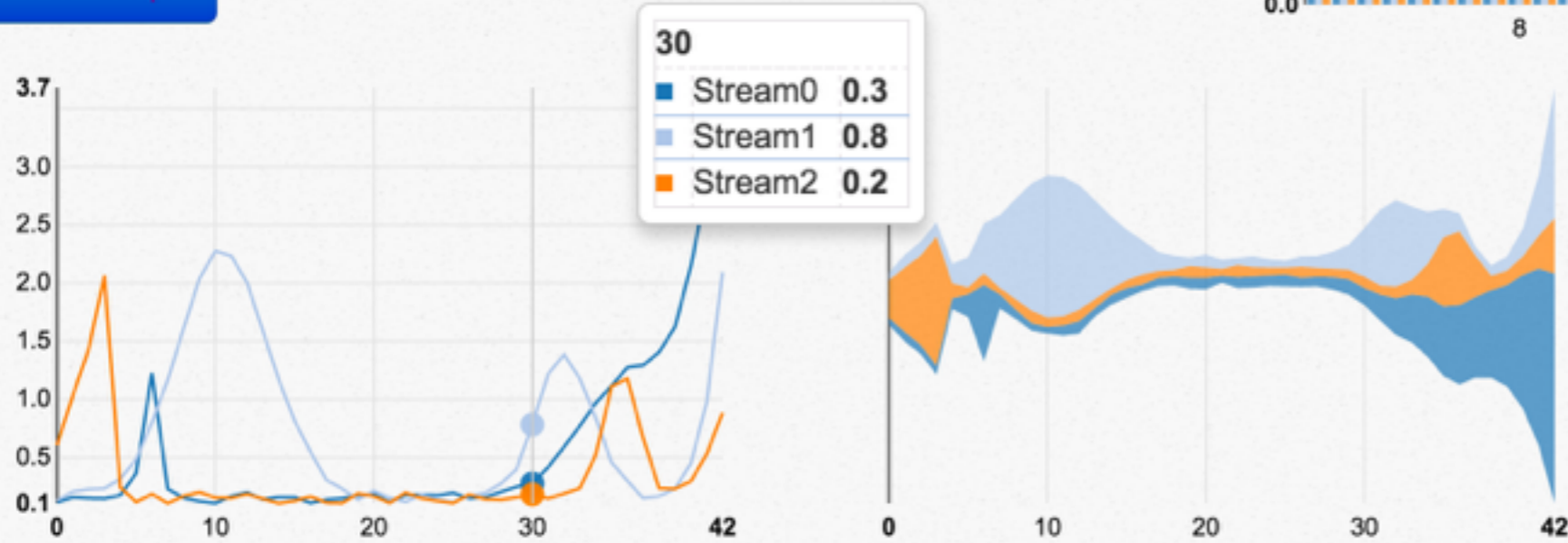
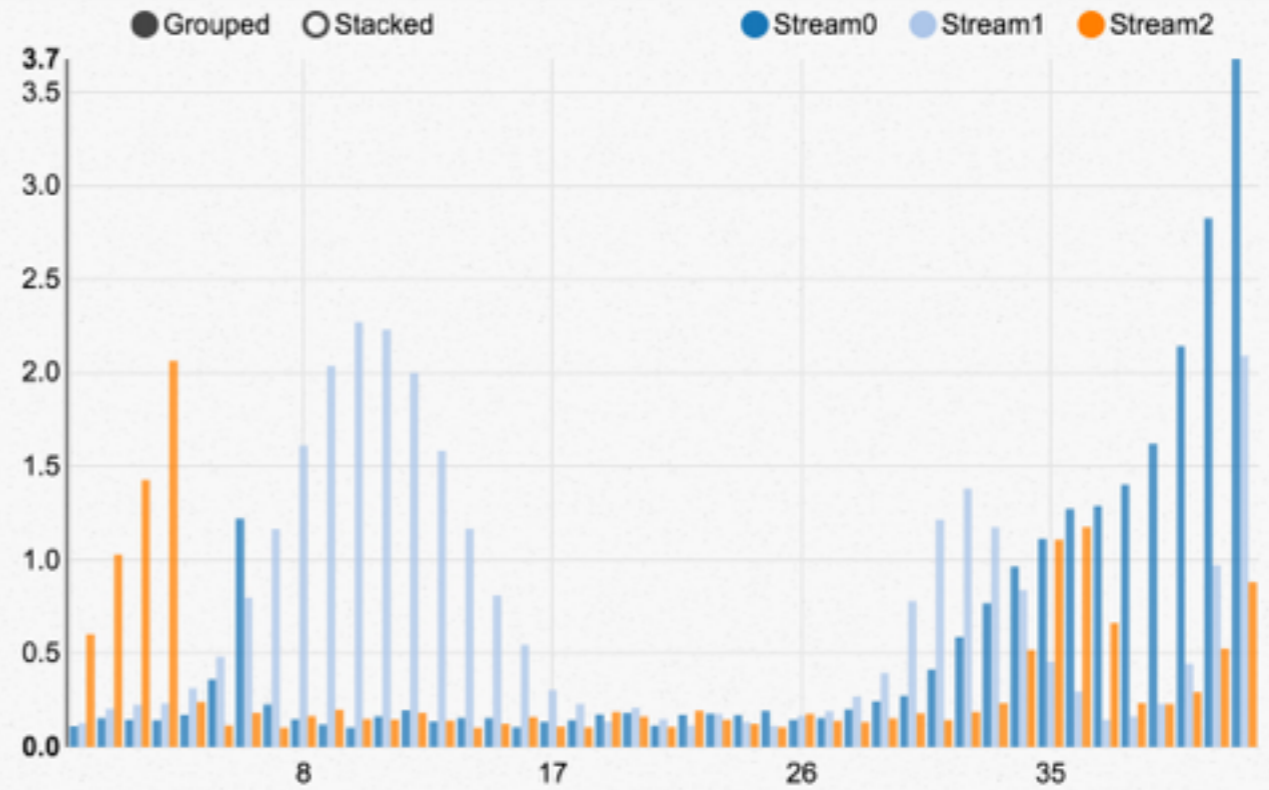
Where do ziz colorz go? O_O

NVD3 Re-usable charts for d3.js

This project is an attempt to build re-usable charts and chart components for [d3.js](#) without taking away the power that [d3.js](#) gives you. This is a very young collection of components, with the goal of keeping these components very customizable, staying away from your standard cookie cutter solutions.

[View more examples »](#)

[GitHub Repo](#)



BREW COLOURS

Number of data classes: ⓘ how to use | updates | downloads | credits

COLORBREWER 2.0
color advice for cartography

Nature of your data: ⓘ
 sequential diverging qualitative

Pick a color scheme:

Multi-hue:

Single hue:

Only show: ⓘ
 colorblind safe
 print friendly
 photocopy safe

Context: ⓘ
 roads
 cities
 borders

Background: ⓘ
 solid color terrain

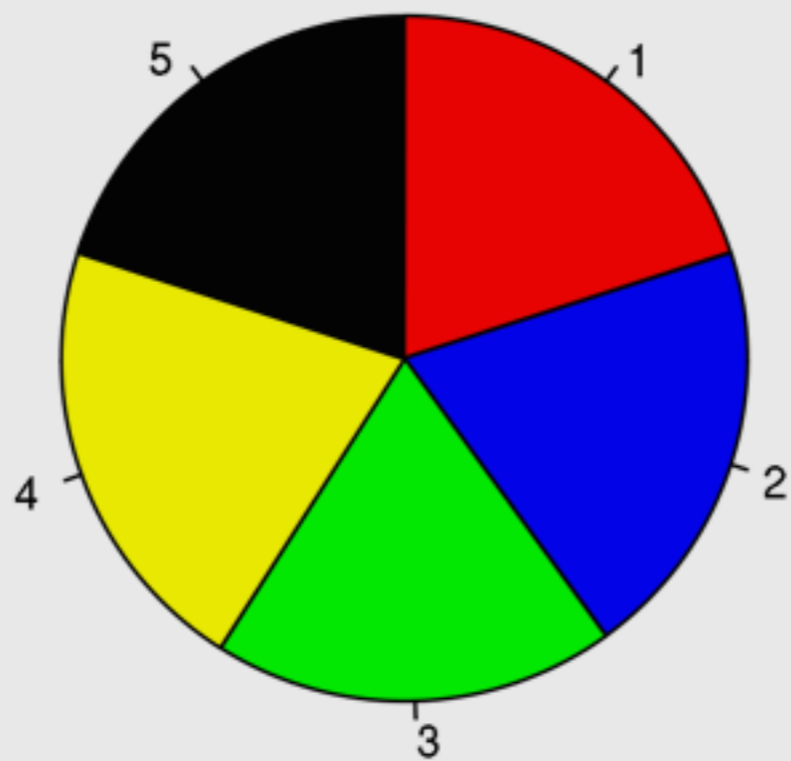
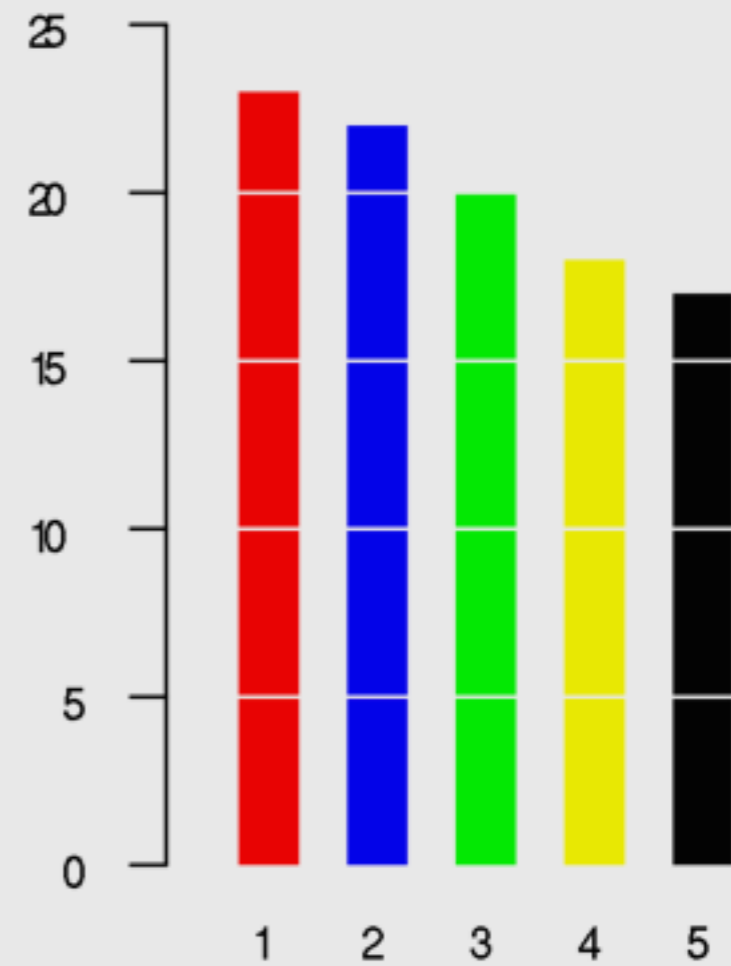
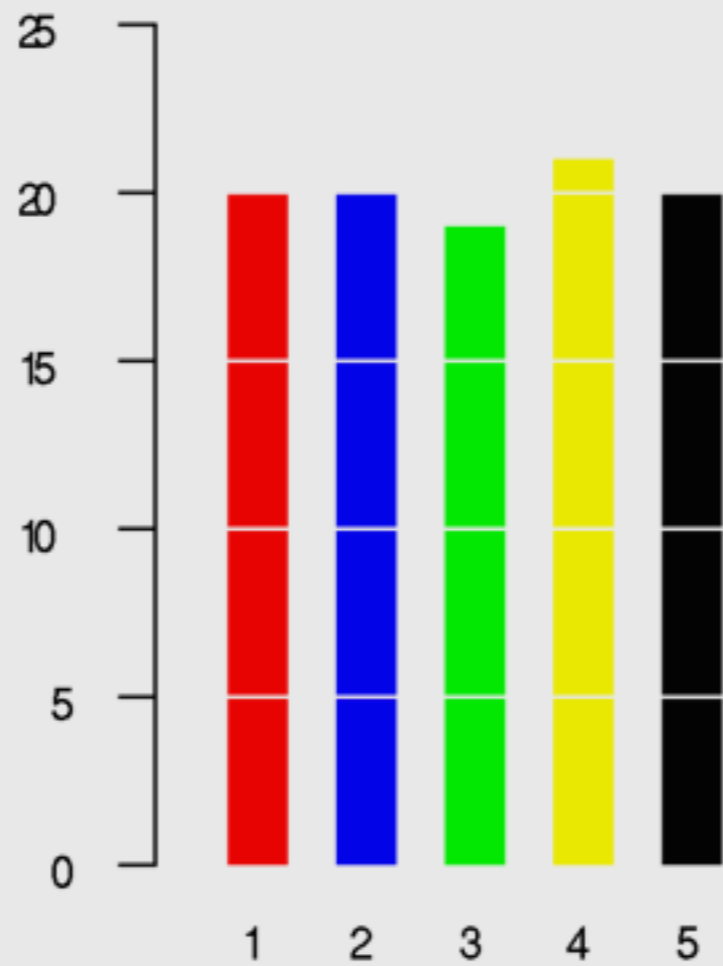
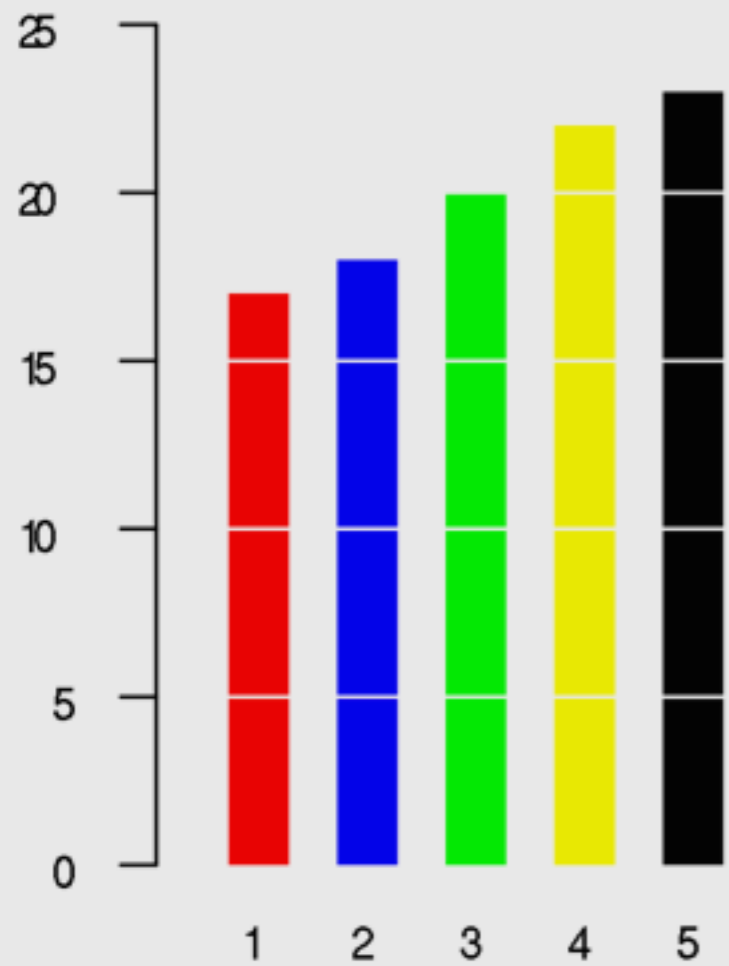
color transparency

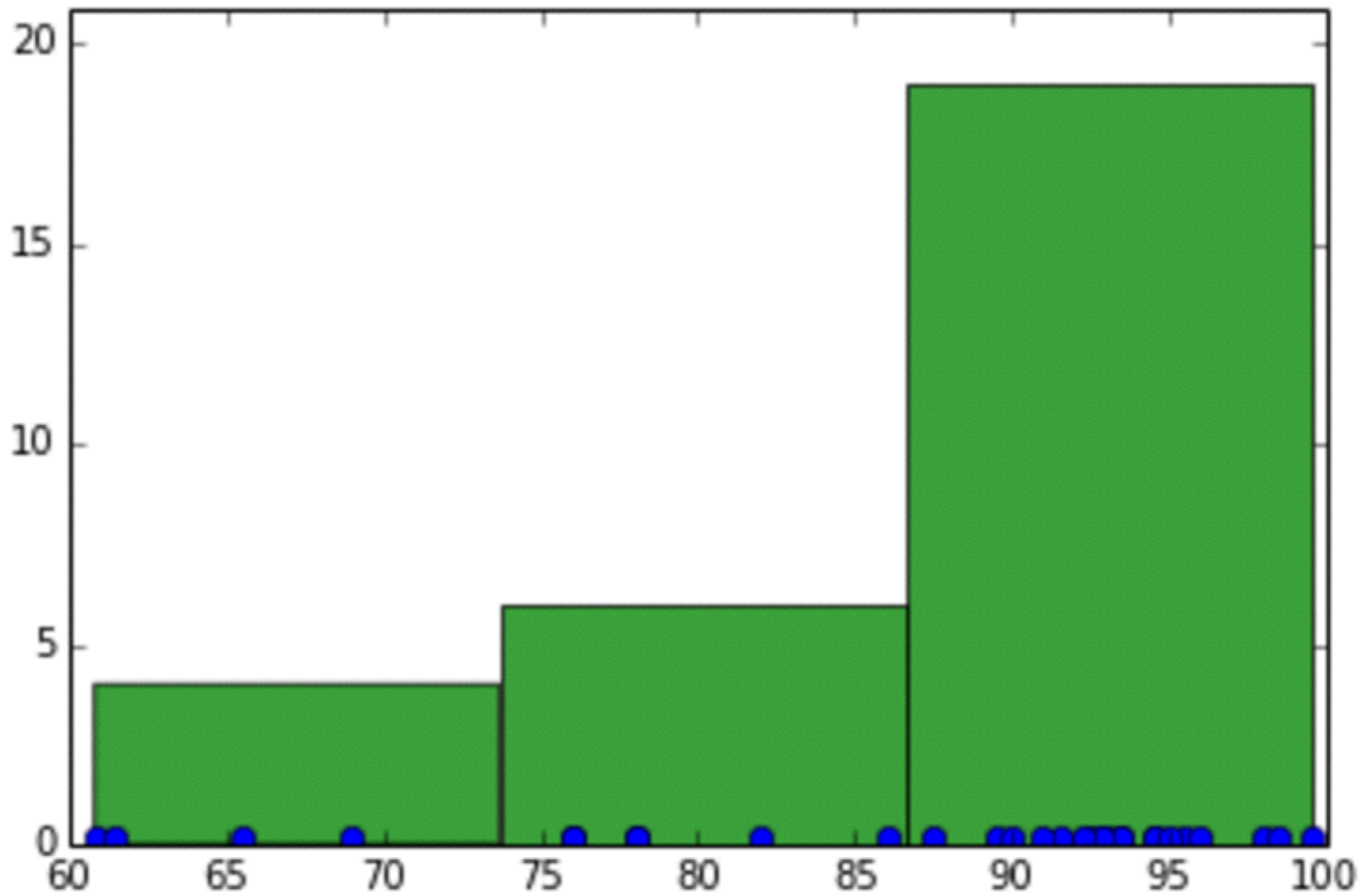
7-class PuBuGn

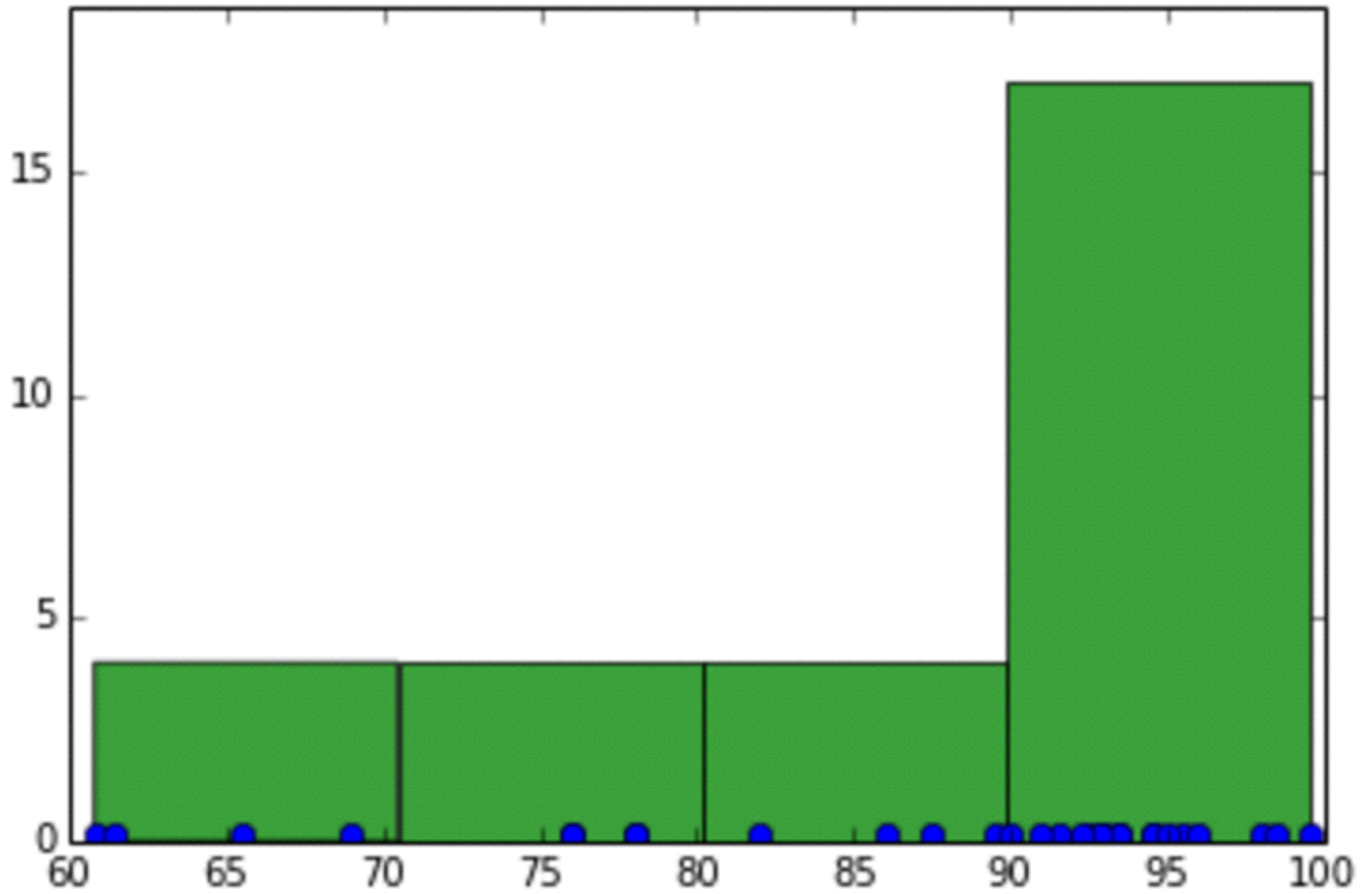
EXPORT

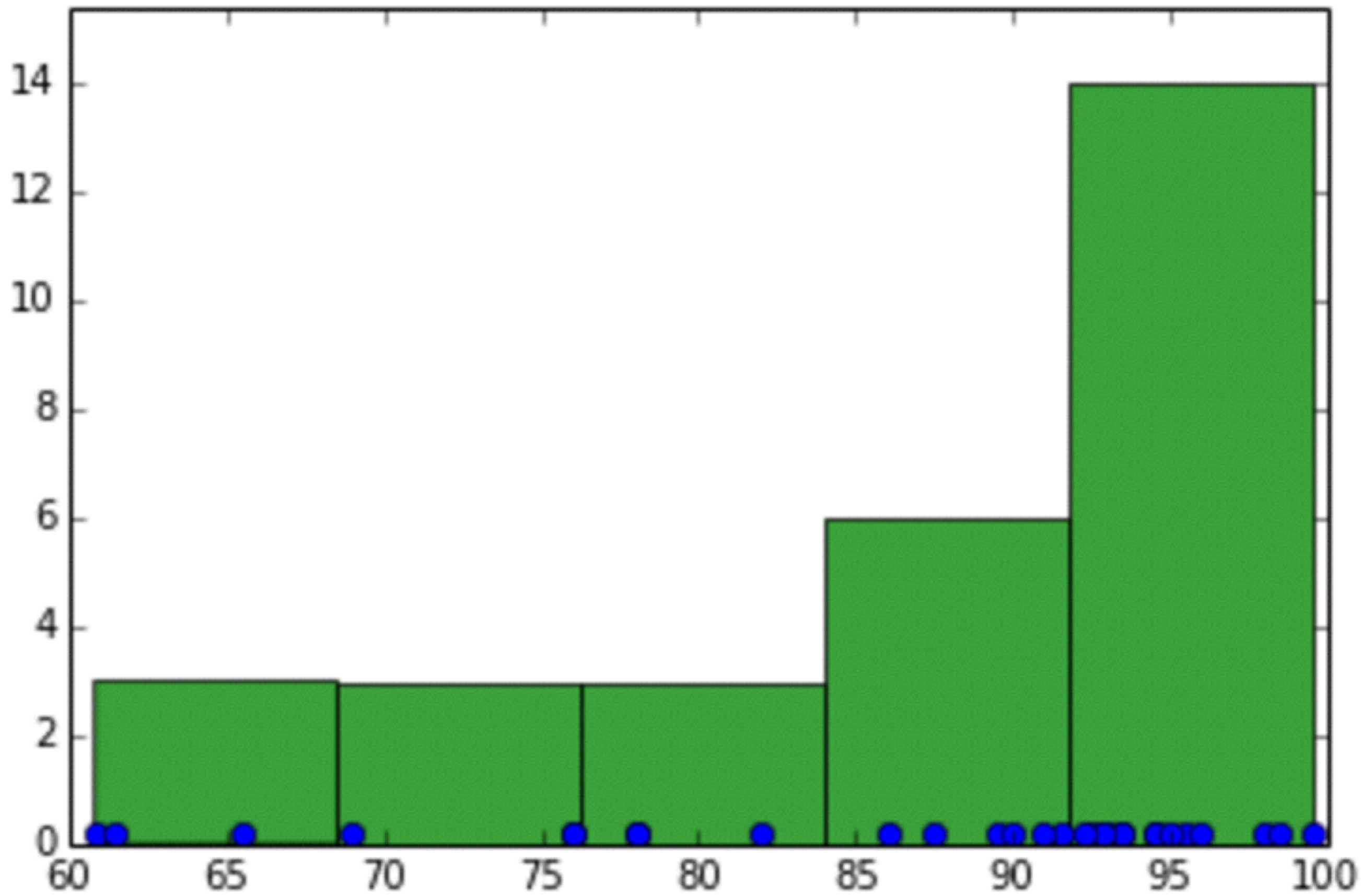
HEX

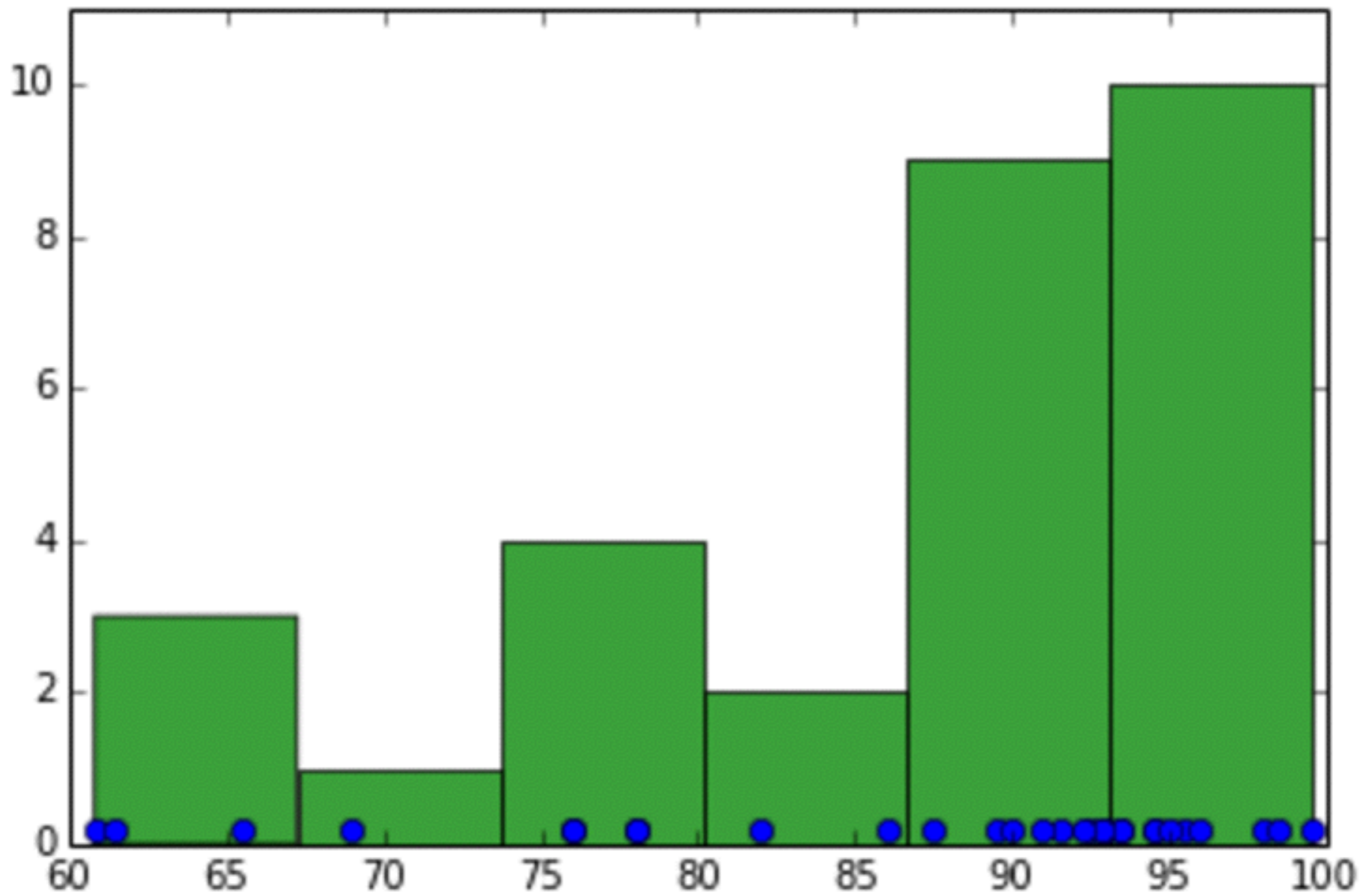
- #f6eff7
- #d0d1e6
- #a6bddb
- #67a9cf
- #3690c0
- #02818a
- #016450

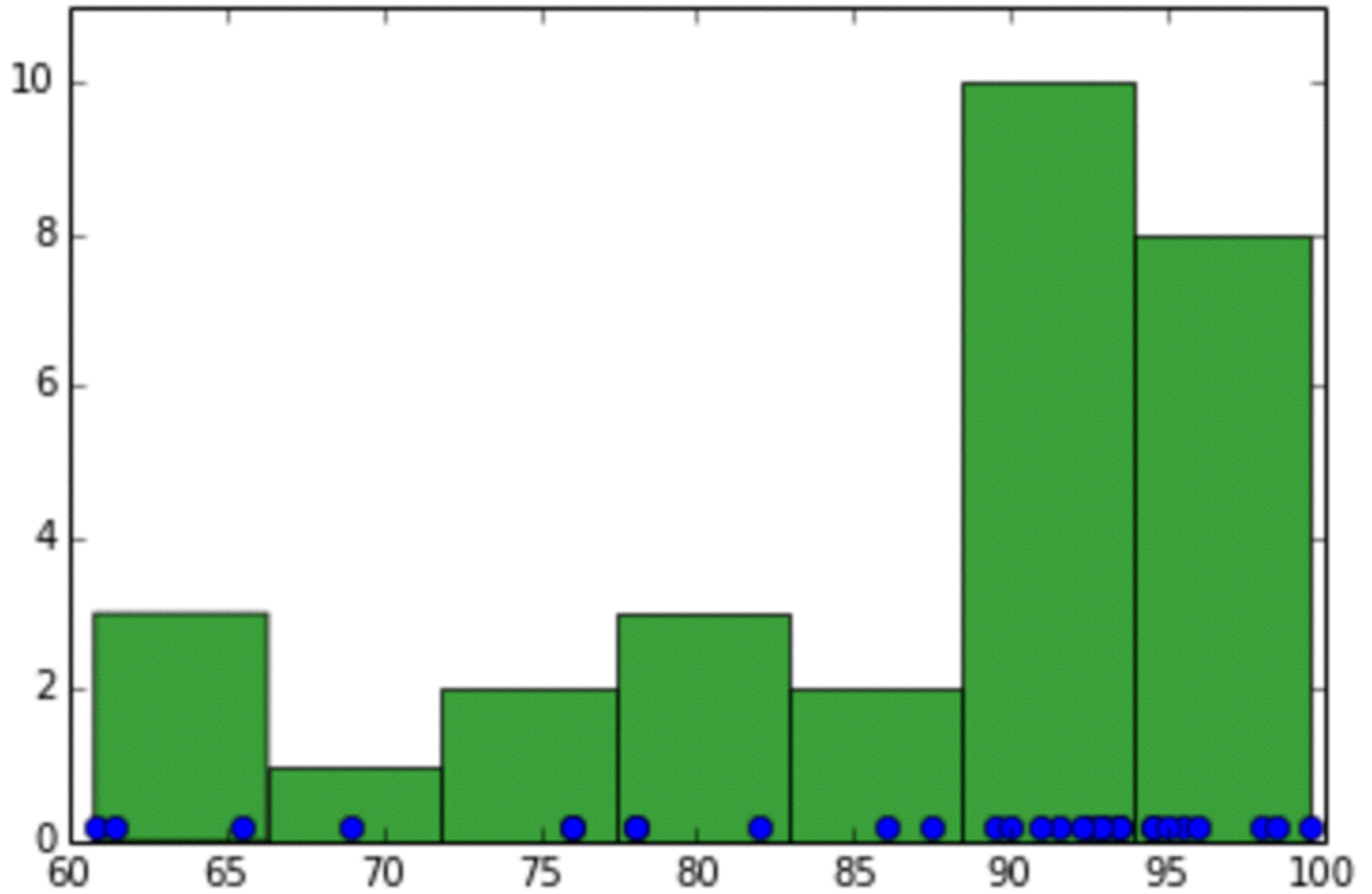
A**B****C**

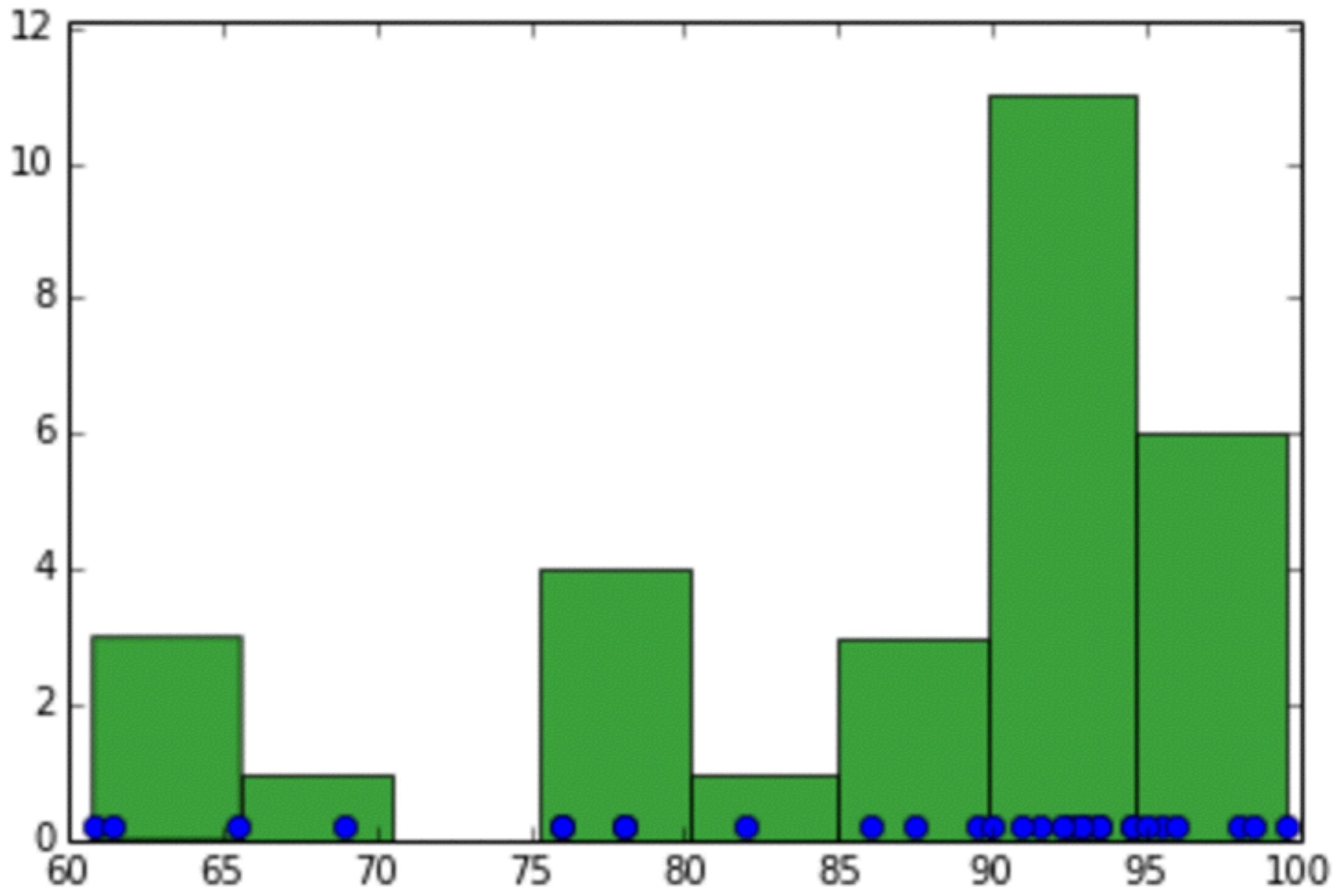












EXISTING TOOLS

SUMMARY

Code history is useful (and fun)

Data cleansing

DIY (by trying simple things)

THANK YOU.

@dmitrykandalov

dmitry.kandalov@gmail.com

<http://github.com/dkandalov>

