

StackTraces – Incidents

R Analysis document

Boris Baldassari – Castalia Solutions

Contents

Introduction	1
About this dataset	1
Terminology	2
Privacy concerns	2
About this document	2
Structure of data	2
All incidents (JSON)	3
Incidents extract (CSV)	4
Bundles extract (CSV)	4
Attributes	5
Message	5
Code	5
Severity	5
Kind	5
Plugin ID	6
Plugin version	9
Status fingerprint	10
Incident fingerprint	10
Incident fingerprint2	10
Timestamp	10
Saved On	11
OSGi Architecture	12
OSGi OS	13
OSGi OS Version	14
OSGi Window Manager	17
Eclipse Build ID	17
Eclipse Product	19
Java runtime version	20
Comment Quality	21
Using the dataset	21
Reading CSV file	21
Using time series (xts)	22
Plot time series	22

Introduction

About this dataset

The Automated Error Reporting (AERI) system retrieves information about exceptions. It is installed by default in the Eclipse IDE and has helped hundreds of projects better support their users and resolve bugs.

This dataset is a dump of all records over a couple of years, with useful information about the exceptions and environment.

- **Generated date:** Fri Feb 23 23:09:29 2018
- **First date:** 2016-03-13 05:00:15
- **Last date:** 2016-12-13 15:49:45
- **Number of incidents:** 2083979
- **Number of attributes:** 20

Terminology

- **Incidents** When an exception occurs and is trapped by the AERI system, it constitutes an incident (or error report). An incident can be reported by several different people, can be reported multiple times, and can be linked to different environments.
- **Problems** As soon as an error report arrives on the server, it will be analyzed and subsequently assigned to one or more problems. A problem thus represents a set of (similar) error reports which usually have the same root cause – for example a bug in your software. (Extract from the AERI system documentation)

This dataset targets only the Incidents of the AERI dataset. There is another dedicated document for the Problems.

Privacy concerns

We value privacy and intend to make everything we can to prevent misuse of the dataset. If you think we failed somewhere in the process, please let us know so we can do better.

The AERI system itself doesn't gather much private information, and takes a great care of it. This dataset goes a step further and removes all identifiable information.

- There is **no email address** in this dataset, **nor any UUID**.
- People not willing to share their traces to the AERI system can tick the private option. This choice has been respected, and all classes that do not belong to public hierarchy have been hidden thanks to an anonymisation mechanism.

The anonymisation technique used basically encrypts information and then throws away the private key. Please refer to the documentation published on github for more details.

About this document

This document is a R Markdown document and is composed of both text (like this one) and dynamically computed information (mostly in the Analysis section below) executed on the data itself. This ensures that the information is always synchronised with the data, and serves as a test suite for the dataset.

Structure of data

The plugin collects a lot of useful information. We only use a subset of it, as required by research interest and privacy protection concerns.

The Incidents dataset comes in two flavours: **All incidents**, in JSON format, and **incidents extract**, in CSV format. There is also a list of bundles discovered in the data dump with their version and number of attached incidents.

All incidents (JSON)

All incidents is the most complete dataset, with all attributes, stacktraces and bundles. Since the stacktraces and bundles structures are too complex for CSV, only the JSON export contains them. The dataset comes as a quite large compressed archive, with one JSON file per incident This represents a total of 2083979 files (incidents).

The structure of an incident file is exemplified below:

```
{
  "eclipseBuildId": "4.6.1.M20160907-1200",
  "eclipseProduct": "org.eclipse.epp.package.jee.product",
  "fingerprint": "cd03d068798d141412b1d1605892fbec",
  "fingerprint2": "12166d864efb7adcccc187034deb7dbf",
  "javaRuntimeVersion": "1.8.0_112-b15",
  "kind": "NORMAL",
  "osgiArch": "x86_64",
  "osgiOs": "Windows7",
  "osgiOsVersion": "6.1.0",
  "osgiWs": "win32",
  "presentBundles": [
    [ "bundle" ]
  ],
  "savedOn": "2016-11-08T10:23:01.914Z",
  "severity": "UNKNOWN",
  "stacktraces": [
    [ "stacktrace" ]
  ],
  "status": {
    "code": 0,
    "fingerprint": "98631af2ddb2d197ebdca532f19d082b",
    "message": "Failed to retrieve default libraries for C:\\\\Program Files\\Java\\jre1.8.0_111",
    "pluginId": "org.eclipse.jdt.launching",
    "pluginVersion": "3.8.100.v20160505-0636",
    "severity": 4
  },
  "timestamp": "2016-11-08T10:22:59.204Z"
}
```

The structure used in the mongodb for stacktraces has been kept as is: it is composed of fields with all information relevant to each line of the stacktrace. Each stacktrace is an array of objects as shown below:

```
[
  {
    "cN": "sun.net.www.http.HttpClient",
    "mN": "parseHTTPHeader",
    "fN": "HttpClient.java",
    "lN": 786,
  }
]
```

Bundles have the following format:

```
{
  "name": "org.eclipse.egit.core",
  "version": "4.1.1.201511131810-r"
}
```

Incidents extract (CSV)

The **Incidents extract** CSV dataset provides the same information as the full JSON dataset, excluding complex structures that cannot be easily formatted in CSV: stacktraces, bundles, products.

Attributes are: `kind`, `severity`, `pluginId`, `pluginVersion`, `code`, `statusFingerprint`, `message`, `statusSeverity`, `fingerprint`, `fingerprint2`, `timestamp`, `savedOn`, `osgiArch`, `osgiOs`, `osgiOsVersion`, `osgiWs`, `eclipseBuildId`, `eclipseProduct`, `javaRuntimeVersion`, `commentQuality`.

Examples are provided at the end of this file to demonstrate how to use it in R.

Bundles extract (CSV)

The **Bundles extract** CSV dataset lists the Eclipse bundles and versions associated to incidents, with the number of incidents for each pair.

bundle_name	bundle_version	value
org.yocto.bc.ui	1.4.0.201304182148	3
org.yocto.bc.ui	1.4.0.201504302002	2
com.ensoftcorp.atlas.ui	3.0.9.201607151215	4
com.ensoftcorp.atlas.ui	3.0.11.201608301612	2
com.ensoftcorp.atlas.ui	3.0.10.201608051140	1
com.ensoftcorp.atlas.ui	3.0.12.201610111322	2
com.ensoftcorp.atlas.ui	2.7.3.201602181733	2
com.sap.ndb.studio.comlog.ui	2.3.5	2
org.eclipse.oomph.predicates	1.6.0.v20161019-0656	2
org.eclipse.oomph.predicates	1.4.0.v20160426-0423	13
org.eclipse.oomph.predicates	1.6.0.v20160920-0456	1
org.eclipse.oomph.predicates	1.3.0.v20150923-0945	3
org.eclipse.oomph.predicates	1.4.0.v20160323-1057	2
org.eclipse.oomph.predicates	1.2.0.v20150915-0926	1
org.eclipse.oomph.predicates	1.3.0.v20160213-0953	18
org.eclipse.oomph.predicates	1.5.0.v20160707-0243	3
org.fordiac.monitoring	1.0.0.201611161153	3
org.fordiac.monitoring	1.0.0.qualifier	3
org.fordiac.monitoring	1.0.0.201608101255	7
org.fordiac.monitoring	1.0.0.201610120928	4
no.sintef.bvr.ui.editor.common	1.0.0.201604201305	8
com.codeaffine.console.core	1.3.0.20160417-1430	1
com.codeaffine.console.core	1.3.0.20160323-1101	4
com.codeaffine.console.core	1.3.0.20160523-1103	1
com.codeaffine.console.core	1.3.0.20161014-1024	2
org.jboss.tools.jmx.jolokia	1.8.2.v20161110-1525	2
org.eclipse.emf.facet.infra.browser	0.3.2.201401021956	2
org.apache.ivyde.eclipse.resolvevisualizer	2.2.0.final-201311091524-RELEASE	2
jp.sf.amateras.stepcounter	3.0.3.201404150124	2
org.jboss.tools.seam.core	3.8.1.v20160607-2131	1

Attributes

Message

- Description: A short text summarising the error.
- Type: String

Code

- Description: The numeric status code logged with the error.
- Type: Integer

Statistical summary:

- Range [-364274674 : 123130]
- 1st Quartile 0
- Median 2
- Mean -1631.272
- 3rd Quartile 4

Severity

- Description: An estimate by the user reporting the error about its perceived severity.
- Type: Factors

Distribution:

- CRITICAL 2025
- MAJOR 1748
- MINOR 1723
- NO_BUG 632
- TRIVIAL 172
- UNKNOWN 2077624

Kind

- Description: The type of error recorded, as identified by the AERI system.
- Type: Factors

The possible values found in the dataset for this attributes are:

- NORMAL (count: 1685795)
- THIRD (count: 357495)
- FREEZE (count: 21612)
- OOME (count: 8731)
- SOE (count: 6237)
- THIRD_FREEZE (count: 4109)

Notes

There are different kinds of incidents described in the official documentation:

- Normal Error: Normal errors are all exceptions that were reported by a client but that are not of kind defined below. Common examples of a normal error are a `NullPointerException` or `IllegalArgumentException`.

- An `OutOfMemoryError` is a special kind of exception. Unlike for normal errors, the stack frame (implicitly) throwing the exception is only sometimes indicative of the root cause of the problem.
- A `StackOverflowError` is a special kind of exception, whose unique characteristic is a repeating pattern of stack frames near the top of the stack trace.
- UI Freeze: A UI freeze is caused by a long-running operation or even a deadlock on the UI thread.
- Third-Party Error: Third-party errors are reports that were received by the Codetrails Error Analytics Server, which deemed neither the configured projects nor their dependencies at fault.
- Third-Party UI Freeze: Third-Party UI Freezes are UI freezes that were received by the Codetrails Error Analytics Server, which deemed neither the configured projects nor their dependencies at fault.

Plugin ID

- Description: The ID of the Eclipse plugin that threw the exception.
- Type: Factors

The possible values found in the dataset for this attributes are:

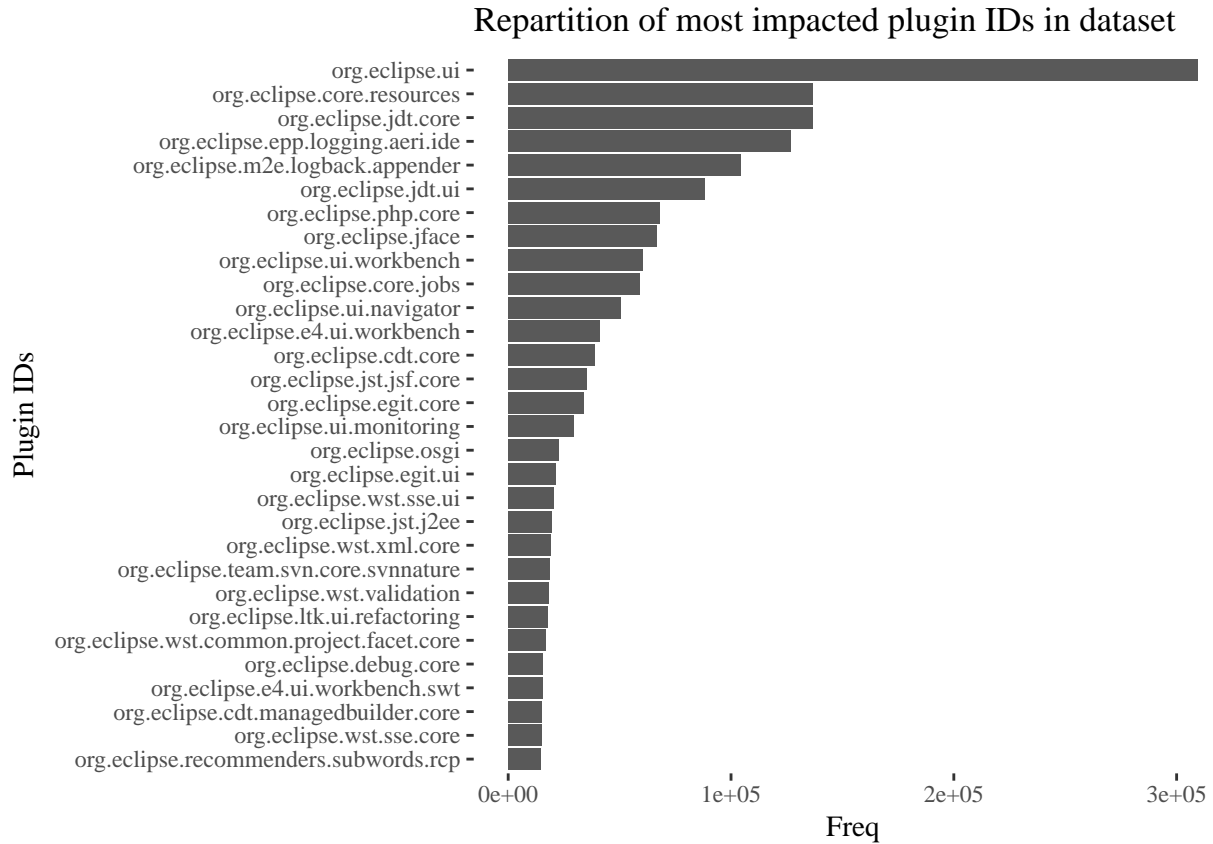
- org.eclipse.ui (count: 309544)
- org.eclipse.core.resources (count: 136748)
- org.eclipse.jdt.core (count: 136471)
- org.eclipse.epp.logging.aeri.ide (count: 126791)
- org.eclipse.m2e.logback.appender (count: 104477)
- org.eclipse.jdt.ui (count: 87982)
- org.eclipse.php.core (count: 68090)
- org.eclipse.jface (count: 66473)
- org.eclipse.ui.workbench (count: 60477)
- org.eclipse.core.jobs (count: 59099)
- org.eclipse.ui.navigator (count: 50409)
- org.eclipse.e4.ui.workbench (count: 41276)
- org.eclipse.cdt.core (count: 38682)
- org.eclipse.jst.jsf.core (count: 35020)
- org.eclipse.egit.core (count: 33742)
- org.eclipse.ui.monitoring (count: 29615)
- org.eclipse.osgi (count: 22558)
- org.eclipse.egit.ui (count: 21537)
- org.eclipse.wst.sse.ui (count: 20599)
- org.eclipse.jst.j2ee (count: 19421)
- org.eclipse.wst.xml.core (count: 18947)
- org.eclipse.team.svn.core.svnnature (count: 18825)
- org.eclipse.wst.validation (count: 18257)
- org.eclipse.ltk.ui.refactoring (count: 17682)
- org.eclipse.wst.common.project.facet.core (count: 16856)
- org.eclipse.debug.core (count: 15644)
- org.eclipse.e4.ui.workbench.swt (count: 15588)
- org.eclipse.cdt.managedbuilder.core (count: 15143)
- org.eclipse.wst.sse.core (count: 15086)
- org.eclipse.recommenders.subwords.rcp (count: 14428)
- org.eclipse.debug.ui (count: 14171)
- org.eclipse.team.core (count: 13835)
- org.eclipse.jdt.debug (count: 13592)
- org.eclipse.andmore (count: 13146)
- org.eclipse.text (count: 12881)
- org.eclipse.dltk.core (count: 11054)
- org.eclipse.core.net (count: 10931)

- org.eclipse.jface.text (count: 10569)
- org.eclipse.epp.mpc.core (count: 10325)
- org.eclipse.equinox.preferences (count: 9496)
- org.eclipse.jst.jsp.core (count: 9359)
- org.eclipse.php.ui (count: 8349)
- org.eclipse.wb.core (count: 8277)
- org.eclipse.search (count: 8047)
- org.eclipse.ant.launching (count: 7767)
- org.eclipse.wst.server.core (count: 7766)
- org.eclipse.jdt.debug.ui (count: 7534)
- org.eclipse.recommenders.rcp (count: 7278)
- org.eclipse.jst.server.tomcat.core (count: 6979)
- org.eclipse.team.cvs.core (count: 6591)
- org.eclipse.jpt.jpa.core (count: 6526)
- org.eclipse.epp.mpc.ui (count: 6524)
- org.eclipse.wst.jsdt.ui (count: 6077)
- org.eclipse.jdt.launching (count: 5708)
- org.eclipse.jst.common.project.facet.core (count: 5638)
- org.eclipse.m2e.core (count: 5614)
- org.eclipse.cdt.ui (count: 5501)
- org.eclipse.core.contenttype (count: 5328)
- org.eclipse.wst.jsdt.core (count: 5066)
- org.eclipse.recommenders.completion.rcp (count: 5059)
- org.eclipse.equinox.registry (count: 4718)
- org.eclipse.team.cvs.ui (count: 4669)
- org.eclipse.oomph.setup.ui (count: 4498)
- org.eclipse.jst.jee (count: 4240)
- org.eclipse.equinox.security (count: 4214)
- org.eclipse.ui.intro (count: 4013)
- org.eclipse.core.filesystem (count: 3929)
- org.eclipse.equinox.p2.operations (count: 3674)
- org.eclipse.core.filebuffers (count: 3669)
- org.eclipse.dltk.core.index.sql (count: 3581)
- org.eclipse.ltk.core.refactoring (count: 3558)
- org.eclipse.datatools.sqltools.sqleditor (count: 3538)
- org.eclipse.equinox.p2.publisher.eclipse (count: 3482)
- org.eclipse.ui.navigator.resources (count: 3396)
- org.eclipse.jst.jsf.facelet.core (count: 3269)
- org.eclipse.rse.ui (count: 3247)
- org.eclipse.e4.ui.workbench.renderers.swt (count: 3104)
- org.eclipse.pde.core (count: 2940)
- org.eclipse.datatools.connectivity (count: 2924)
- org.eclipse.php.debug.core (count: 2900)
- org.eclipse.ui.ide (count: 2857)
- org.eclipse.help.base (count: 2603)
- org.eclipse.buildship.core (count: 2588)
- org.eclipse.cdt.codan.core (count: 2509)
- org.eclipse.jst.j2ee.core (count: 2492)
- org.eclipse.cdt.dsf (count: 2425)
- org.eclipse.core.databinding (count: 2407)
- org.eclipse.core.expressions (count: 2353)
- org.eclipse.jsch.core (count: 2307)
- org.eclipse.jst.servlet.ui (count: 2166)
- org.eclipse.team.ui (count: 2114)

- org.eclipse.mylyn.discovery.core (count: 2111)
- org.eclipse.wst.common.frameworks (count: 1926)
- org.eclipse.cdt.ds.gdb (count: 1857)
- org.eclipse.rap.ui.workbench (count: 1777)
- org.eclipse.help (count: 1661)
- org.eclipse.update.configurator (count: 1586)
- org.eclipse.cdt.launch (count: 1514)
- org.eclipse.core.runtime (count: 1491)
- org.eclipse.jst.jsf.common (count: 1413)
- org.eclipse.wst.common.modulecore (count: 1412)
- org.eclipse.launchbar.core (count: 1411)
- org.eclipse.recommenders.coordinates.rcp (count: 1383)
- org.eclipse.recommenders.types.rcp (count: 1344)
- com.ctrlflow.aer.server-PRODUCTION (count: 1341)
- org.eclipse.ldt (count: 1321)
- org.eclipse.jdt.apt.core (count: 1310)
- org.eclipse.jst.jee.ejb (count: 1297)
- org.eclipse.wst.json.core (count: 1291)
- org.eclipse.equinox.ds (count: 1232)
- org.eclipse.ui.views.properties.tabbed (count: 1232)
- org.eclipse.cdt.codan.core.cxx (count: 1210)
- org.eclipse.wst.xml.ui (count: 1208)
- org.eclipse.dltk.ui (count: 1197)
- org.eclipse.jpt.common.core (count: 1195)
- org.eclipse.ui.intro.universal (count: 1185)
- org.eclipse.mylyn.commons.repositories.core (count: 1173)
- org.eclipse.compare (count: 1104)
- org.eclipse.jdt.junit (count: 1054)
- org.eclipse.php.formatter.core (count: 1040)
- org.eclipse.dltk.core.index.lucene (count: 978)
- org.eclipse.ui.console (count: 963)
- org.eclipse.jst.jee.web (count: 957)
- org.eclipse.vjet.eclipse.core (count: 885)
- org.eclipse.jpt.common.ui (count: 877)
- org.eclipse.ant.core (count: 871)
- org.eclipse.wst.jsdt.web.core (count: 857)
- org.eclipse.equinox.p2.metadata.repository (count: 853)
- org.eclipse.ui.editors (count: 806)
- org.eclipse.m2e.discovery (count: 768)
- org.eclipse.jdt.junit.core (count: 760)
- org.eclipse.graphiti.ui (count: 747)
- org.eclipse.ui.workbench.texteditor (count: 723)
- org.eclipse.papyrus.infra.core (count: 697)
- org.eclipse.andmore.ddms (count: 687)
- org.eclipse.dltk.mod.core (count: 682)
- org.eclipse.rse.core (count: 681)
- org.eclipse.cdt.make.core (count: 669)
- org.eclipse.jdt.apt.pluggable.core (count: 660)
- org.eclipse.sirius (count: 635)
- org.eclipse.oomph.setup.core (count: 625)
- org.eclipse.jst.j2ee.navigator.ui (count: 615)
- org.eclipse.emf.transaction (count: 614)
- org.eclipse.ajdt.core (count: 613)
- org.eclipse.m2e.editor (count: 583)

- org.eclipse.epp.logging.aeri.ui (count: 579)
- org.eclipse.cdt.launch.remote (count: 571)
- org.eclipse.wst.common.emfworkbench.integration (count: 551)
- org.eclipse.gmf.runtime.diagram.ui (count: 538)
- org.eclipse.cft.server.core (count: 525)

Visualisation of the most used Eclipse Build IDs in the dataset:

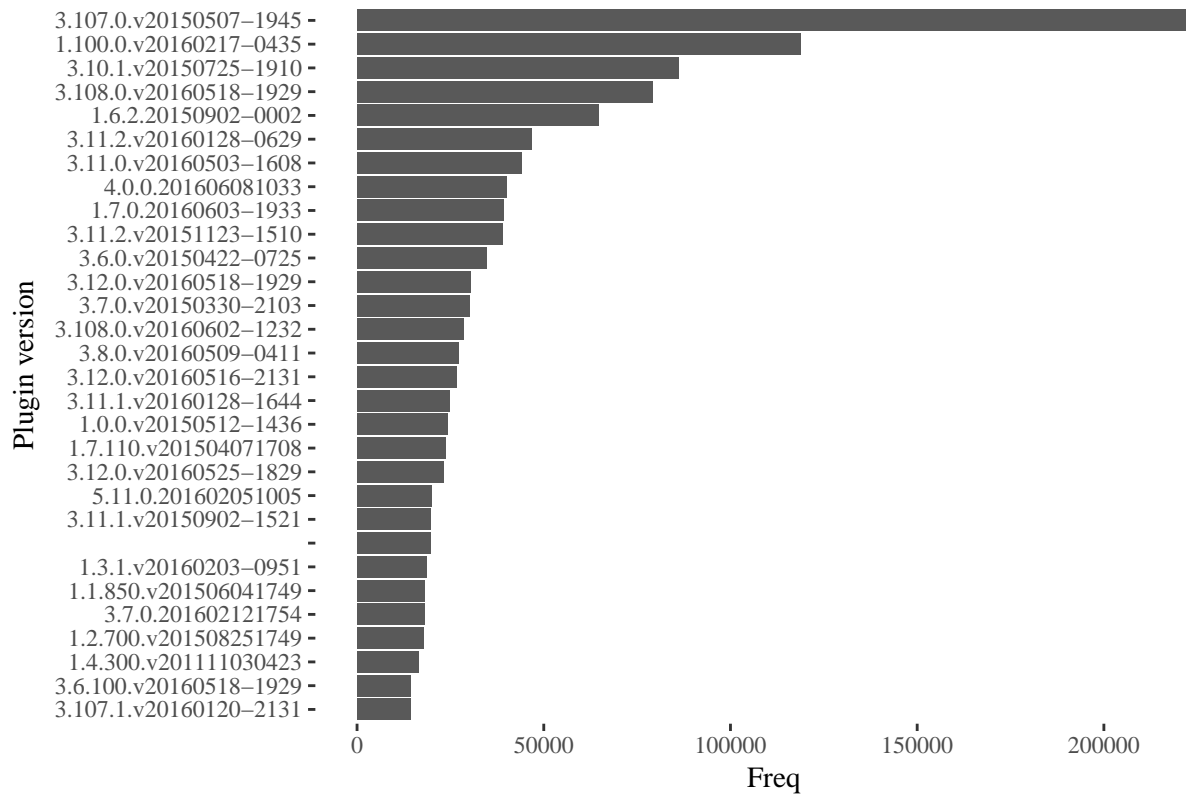


Plugin version

- Description: The ID of the Eclipse plugin that threw the exception.
- Type: Factors

There are 3463 different values found in the dataset for this attribute. The following bar plot only displays the values with more than 500 occurrences:

Repartition of top Eclipse plugin versions in dataset



Status fingerprint

- Description: An identifier for the status of the incident. Used for duplicates detection.
- Type: String

Incident fingerprint

- Description: An identifier for the incident. Used for duplicates detection.
- Type: String

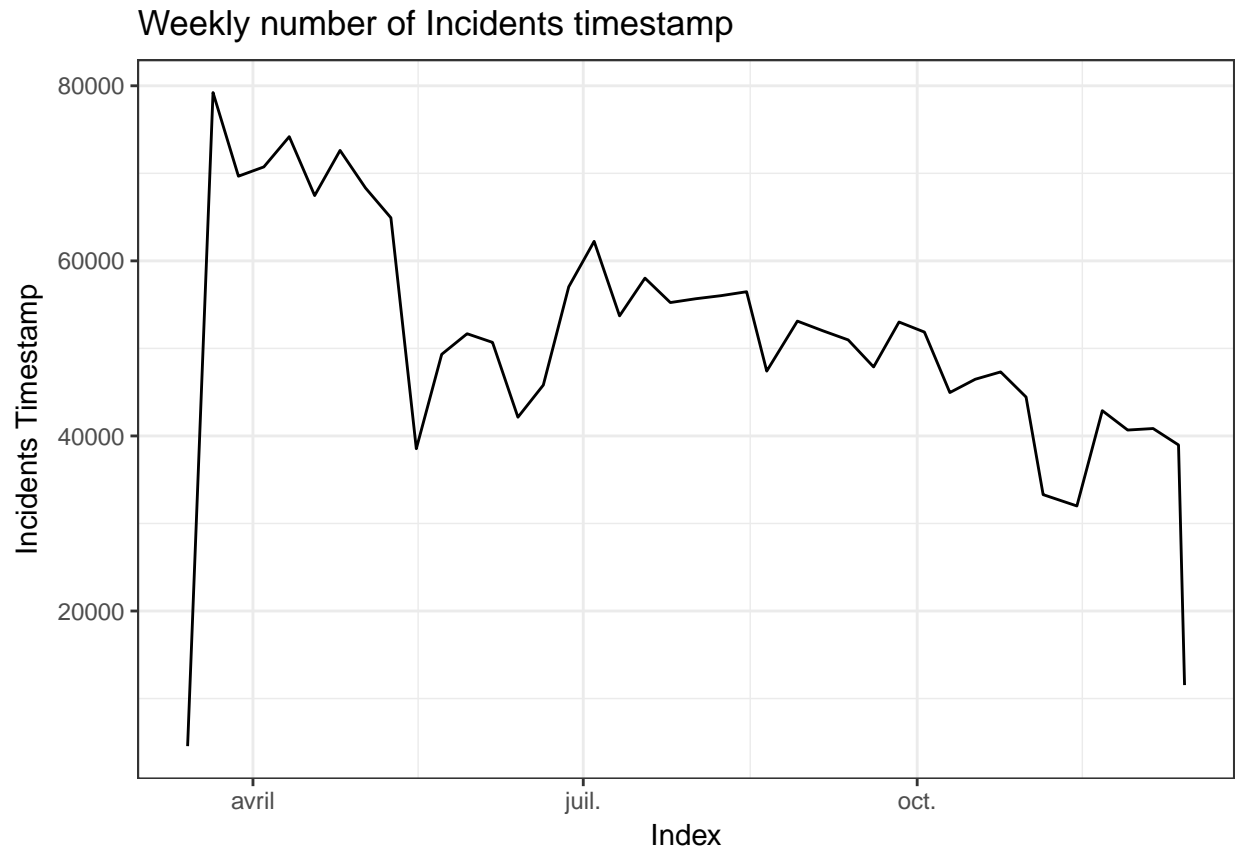
Incident fingerprint2

- Description: An identifier for the incident. Used for duplicates detection.
- Type: String

Timestamp

- Description: The time of creation of the incident.
- Type: Date (ISO8601)

Dates range from 2016-03-13 05:00:15 to 2016-12-13 15:49:45.

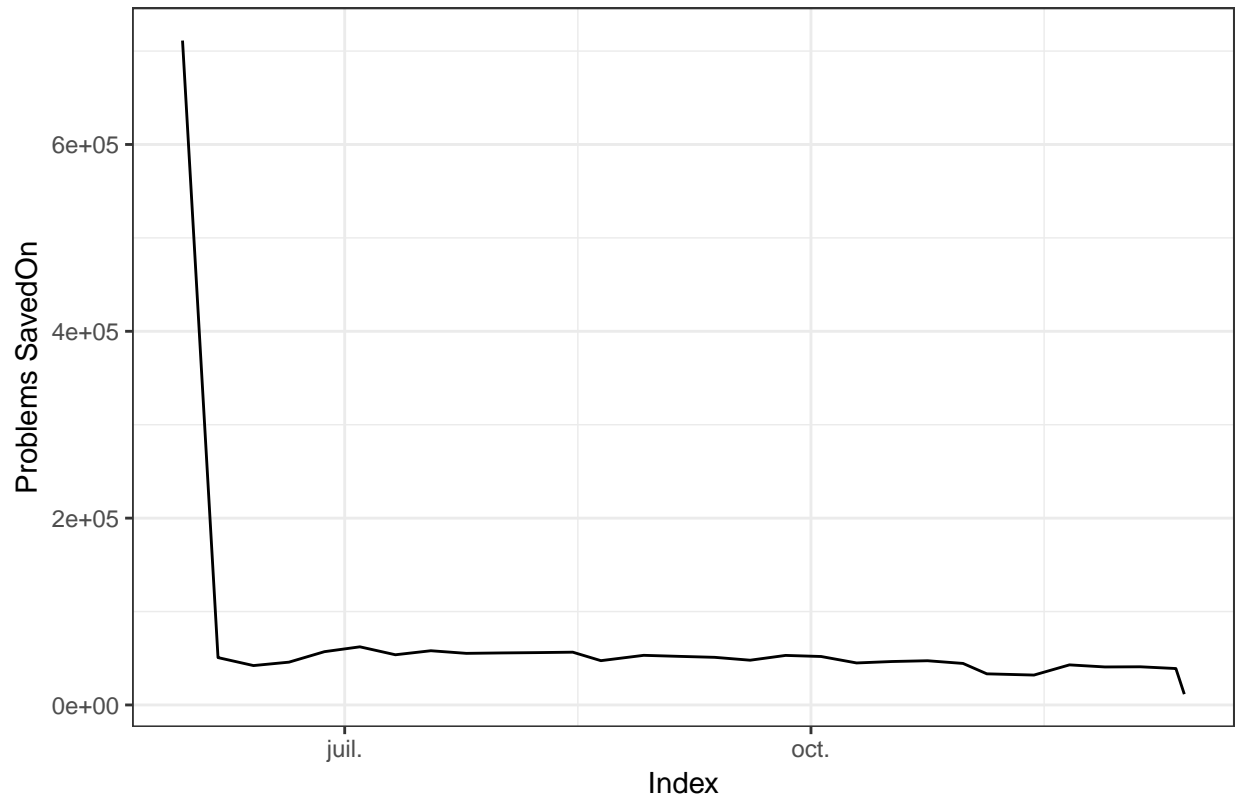


Saved On

- Description: The time of last save of the problem.
- Type: Date (ISO8601)

Dates range from 2016-05-23 05:23:51 to 2016-12-13 15:49:45.

Weekly number of Problems SavedOn



OSGi Architecture

- Description: The architecture of the host, as specified in the OSGi bundle definition.
- Type: Factors

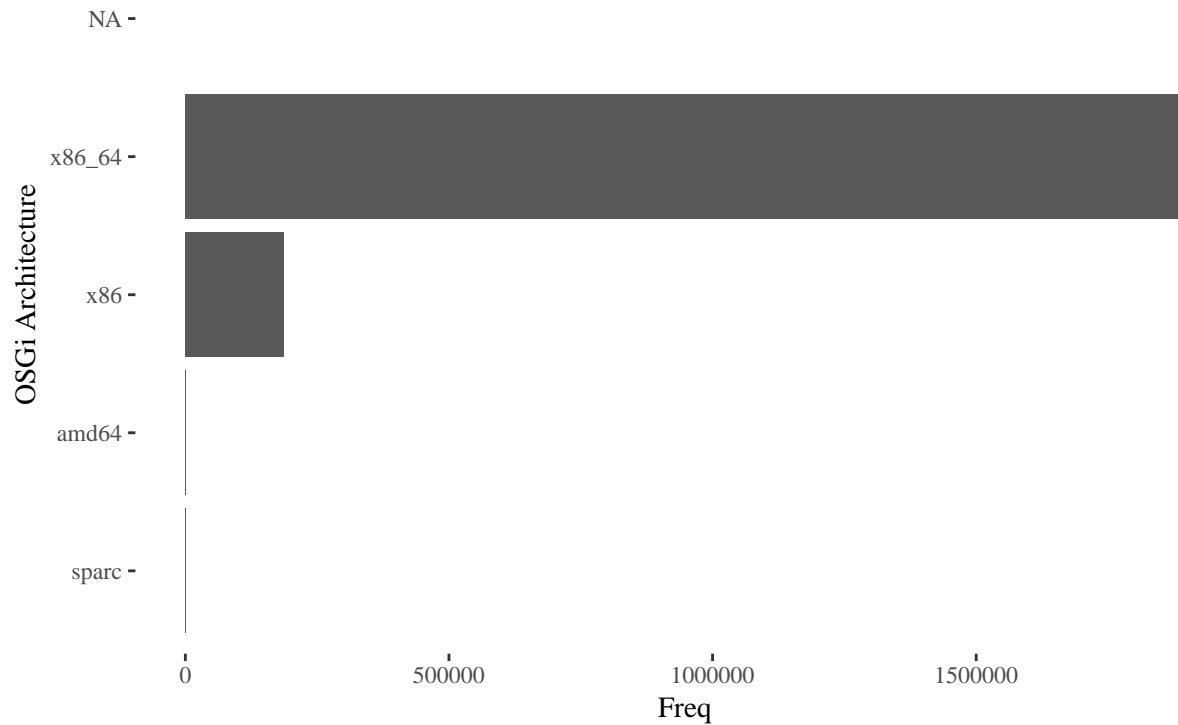
Possible values found in the dataset for this attribute are:

- x86_64 (count: 1894980)
- x86 (count: 187579)
- amd64 (count: 1412)
- sparc (count: 8)

Repartition of architectures:

Warning: Removed 26 rows containing missing values (position_stack).

Repartition of OSGi Architectures in dataset



OSGi OS

- Description: The host operating system, as reported in OSGi.
- Type: Factors

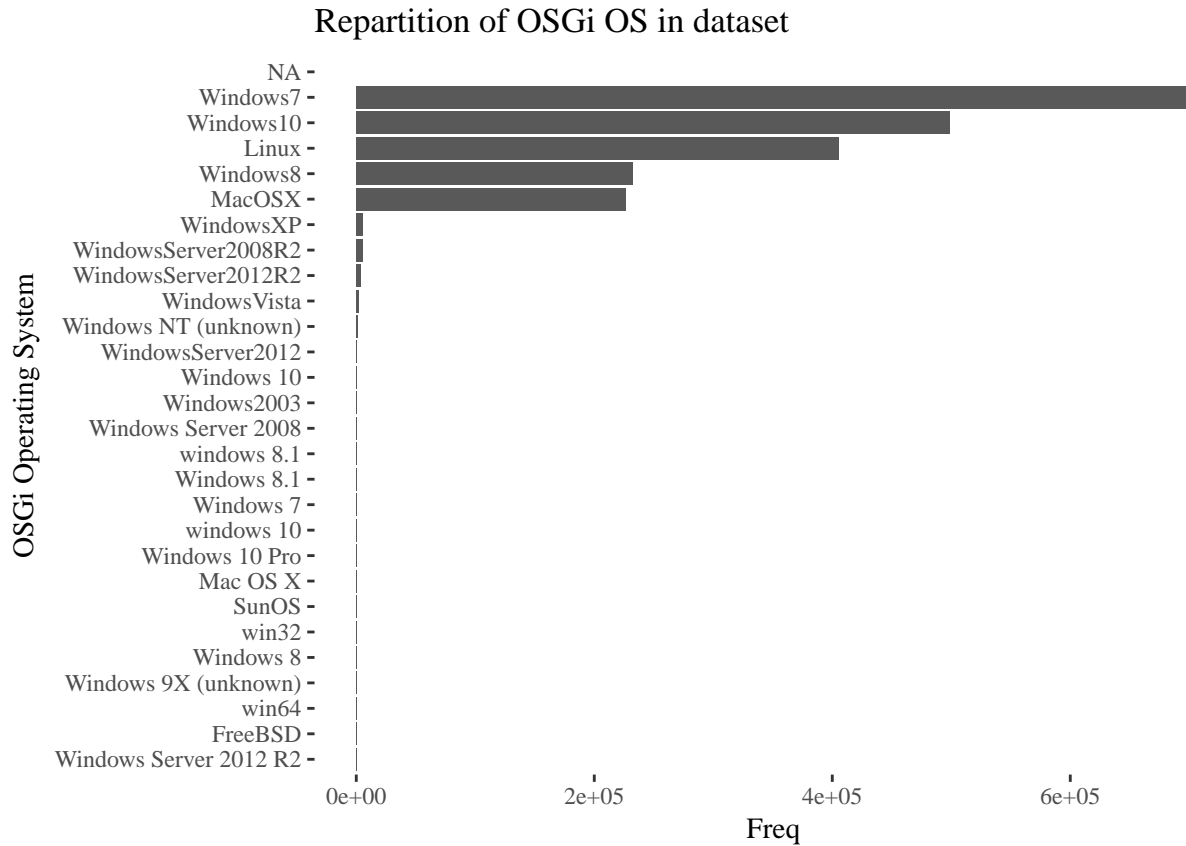
The possible values found in the dataset for this attributes are:

- Windows7 (count: 701169)
- Windows10 (count: 498429)
- Linux (count: 405317)
- Windows8 (count: 232059)
- MacOSX (count: 226714)
- WindowsXP (count: 5677)
- WindowsServer2008R2 (count: 5479)
- WindowsServer2012R2 (count: 3422)
- WindowsVista (count: 2431)
- Windows NT (unknown) (count: 1591)
- WindowsServer2012 (count: 593)
- Windows 10 (count: 349)
- Windows2003 (count: 203)
- Windows Server 2008 (count: 200)
- windows 8.1 (count: 143)
- Windows 7 (count: 42)
- Windows 8.1 (count: 42)
- windows 10 (count: 29)
- Windows 10 Pro (count: 23)

- Mac OS X (count: 17)
- SunOS (count: 16)
- win32 (count: 13)
- Windows 8 (count: 10)
- Windows 9X (unknown) (count: 5)
- win64 (count: 3)
- FreeBSD (count: 2)
- Windows Server 2012 R2 (count: 1)

Visualisation of the various operating systems used in the dataset:

Warning: Removed 3 rows containing missing values (position_stack).



OSGi OS Version

- Description: The host operating system version, as reported in OSGi.
- Type: Factors

The possible values found in the dataset for this attributes are:

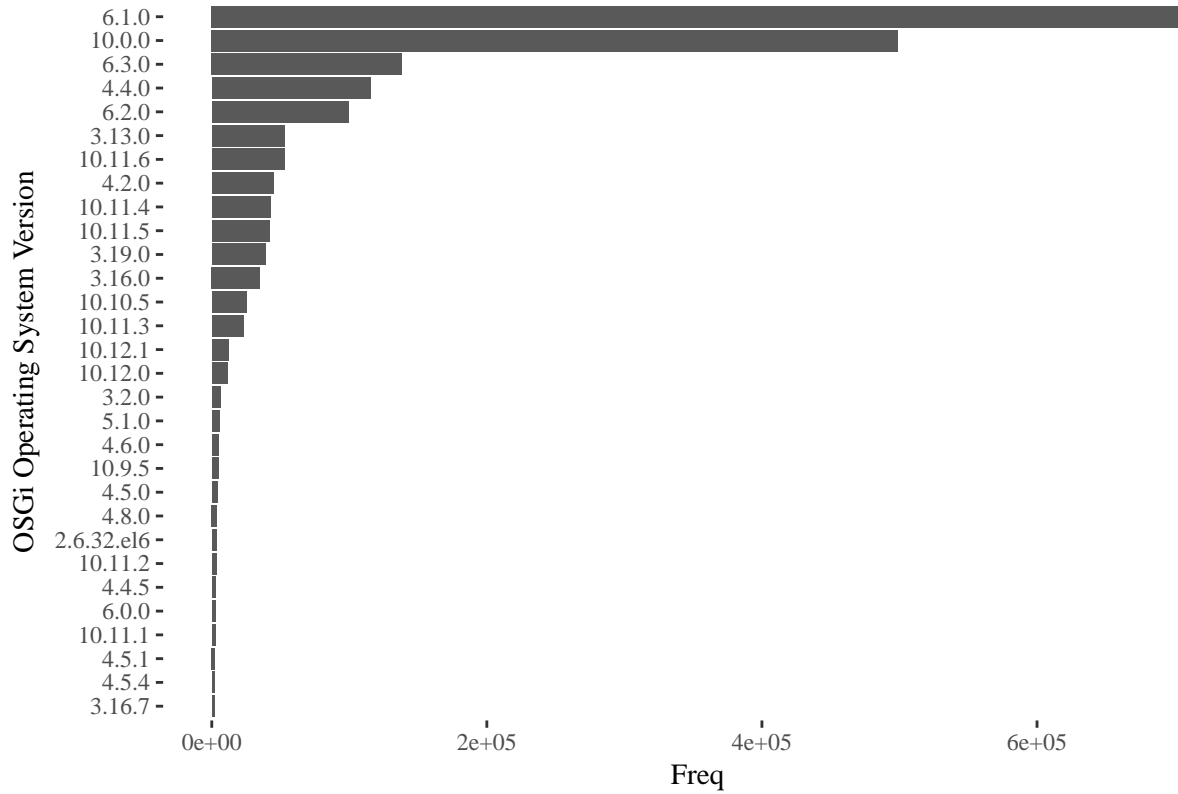
- 6.1.0 (count: 706697)
- 10.0.0 (count: 498962)
- 6.3.0 (count: 138231)
- 4.4.0 (count: 115414)
- 6.2.0 (count: 99513)
- 3.13.0 (count: 52918)
- 10.11.6 (count: 52857)

- 4.2.0 (count: 44710)
- 10.11.4 (count: 42881)
- 10.11.5 (count: 42022)
- 3.19.0 (count: 39154)
- 3.16.0 (count: 35033)
- 10.10.5 (count: 25287)
- 10.11.3 (count: 22944)
- 10.12.1 (count: 12135)
- 10.12.0 (count: 11563)
- 3.2.0 (count: 6285)
- 5.1.0 (count: 5633)
- 4.6.0 (count: 4851)
- 10.9.5 (count: 4729)
- 4.5.0 (count: 4105)
- 4.8.0 (count: 3920)
- 2.6.32.el6 (count: 3565)
- 10.11.2 (count: 3353)
- 4.4.5 (count: 2741)
- 6.0.0 (count: 2629)
- 10.11.1 (count: 2580)
- 4.5.1 (count: 2429)
- 4.5.4 (count: 2108)
- 3.16.7 (count: 2047)
- 3.10.0.36 (count: 1904)
- 4.7.0 (count: 1880)
- 3.10.0.28 (count: 1875)
- 4.6.2 (count: 1719)
- 3.10.0.22 (count: 1683)
- 3.10.0.el7 (count: 1638)
- 3.10.0.10 (count: 1568)
- 4.6.4 (count: 1508)
- 3.10.0.13 (count: 1488)
- 4.4.6 (count: 1460)
- 4.1.15 (count: 1431)
- 4.6.3 (count: 1385)
- 3.10.0.18 (count: 1367)
- 10.11.0 (count: 1334)
- 2.6.32.18 (count: 1324)
- 4.4.8 (count: 1316)
- 2.6.32.22 (count: 1307)
- 4.4.6.fc23 (count: 1287)
- 2.6.32.12 (count: 1203)
- 2.6.32.3 (count: 1177)
- 2.6.32.26 (count: 1117)
- 3.0.101-0.47.71-xen (count: 1114)
- 4.7.2 (count: 1096)
- 2.6.32.1 (count: 1076)
- 3.11.0 (count: 1037)
- 10.10.1 (count: 1003)
- 10.10.3 (count: 982)
- 4.3.0 (count: 969)
- 4.5.7.fc23 (count: 964)
- 4.1.31 (count: 961)
- 4.6.4.fc24 (count: 878)

- 3.8.0 (count: 869)
- 4.1.13.fc21 (count: 860)
- 4.1.27 (count: 839)
- 4.6.7.fc24 (count: 826)
- 4.6.3.fc24 (count: 781)
- 2.6.32.4 (count: 735)
- 2.6.32.11 (count: 728)
- 4.4.9.fc23 (count: 728)
- 3.10.0.4 (count: 721)
- 4.4.2.ign (count: 699)
- 4.7.4 (count: 693)
- 2.6.32.6 (count: 686)
- 3.5.0 (count: 635)
- 10.10.4 (count: 611)
- 4.4.6.fc22 (count: 606)
- 4.1.20 (count: 568)
- 4.7.4.fc24 (count: 568)
- 4.5.5.fc24 (count: 562)
- 4.7.6 (count: 542)
- 4.4.13 (count: 513)
- 10.10.2 (count: 508)

Visualisation of the various operating system versions used in the dataset:

Repartition of most used OSGi OS versions in dataset



OSGi Window Manager

- Description: The Window Manager used by the host, as reported in OSGi.
- Type: Factors

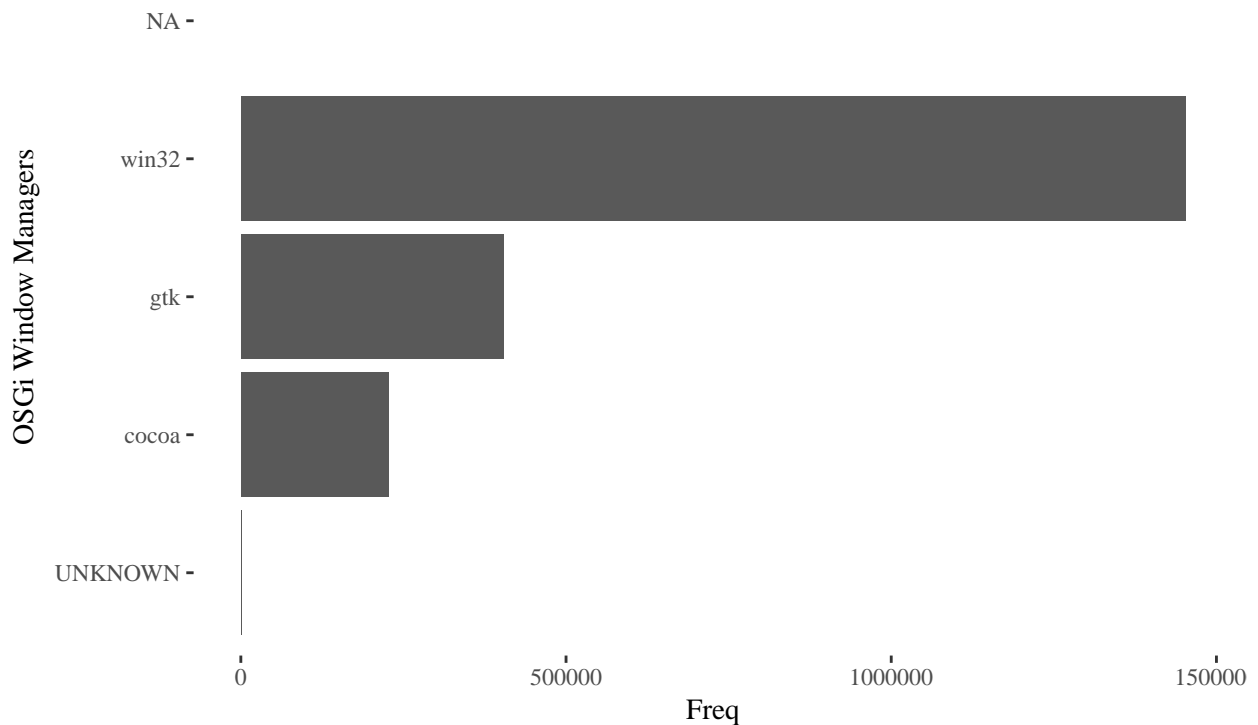
The possible values found in the dataset for this attributes are:

- win32 (count: 1451913)
- gtk (count: 403923)
- cocoa (count: 226714)
- UNKNOWN (count: 1429)

Visualisation of the various Window managers used in the dataset:

Warning: Removed 26 rows containing missing values (position_stack).

Repartition of OSGi Window managers in dataset



Eclipse Build ID

- Description: The Build ID of the Eclipse instance running when the exception occurred.
- Type: Factors

The possible values found in the dataset for this attributes are:

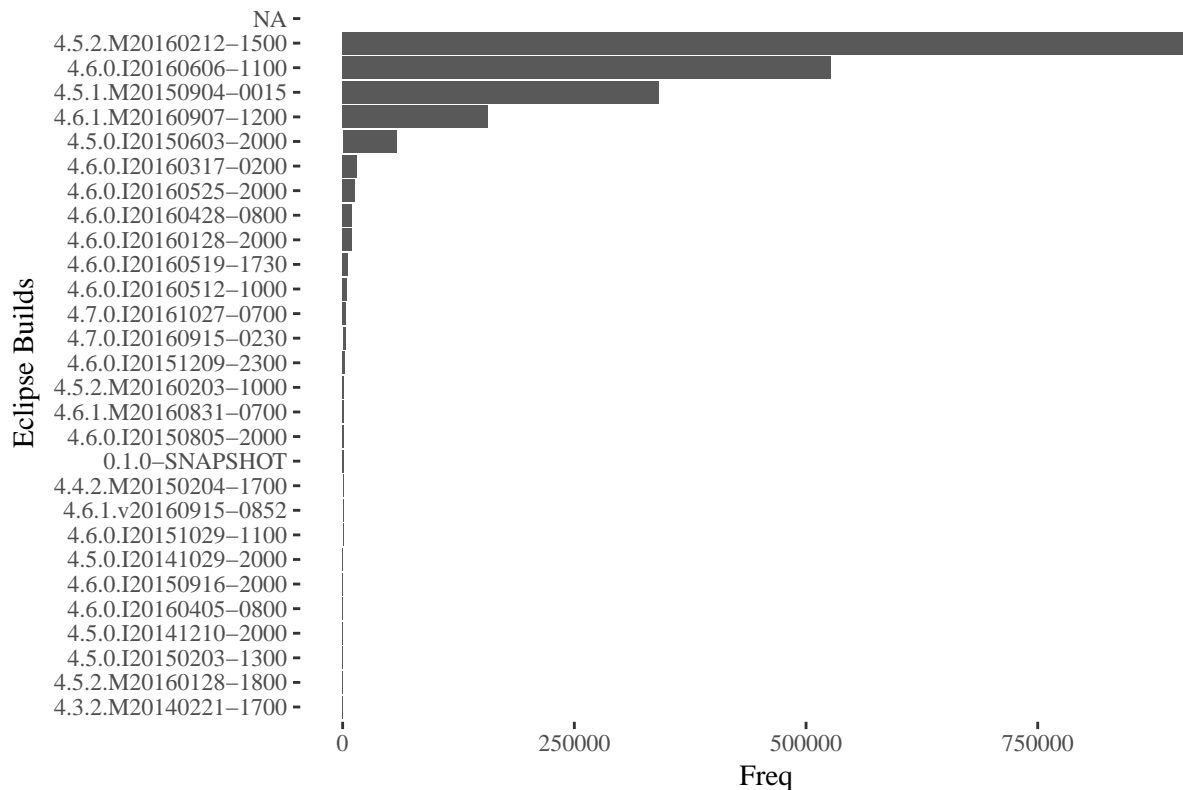
- 4.5.2.M20160212-1500 (count: 914332)
- 4.6.0.I20160606-1100 (count: 527265)
- 4.5.1.M20150904-0015 (count: 341505)
- 4.6.1.M20160907-1200 (count: 156663)
- 4.5.0.I20150603-2000 (count: 58043)

- 4.6.0.I20160317-0200 (count: 15343)
- 4.6.0.I20160525-2000 (count: 13521)
- 4.6.0.I20160428-0800 (count: 10014)
- 4.6.0.I20160128-2000 (count: 9864)
- 4.6.0.I20160519-1730 (count: 5565)
- 4.6.0.I20160512-1000 (count: 4798)
- 4.7.0.I20161027-0700 (count: 3607)
- 4.7.0.I20160915-0230 (count: 3164)
- 4.6.0.I20151209-2300 (count: 2821)
- 4.5.2.M20160203-1000 (count: 1689)
- 4.6.1.M20160831-0700 (count: 1666)
- 4.6.0.I20150805-2000 (count: 1484)
- 0.1.0-SNAPSHOT (count: 1429)
- 4.4.2.M20150204-1700 (count: 999)
- 4.6.1.v20160915-0852 (count: 938)
- 4.6.0.I20151029-1100 (count: 904)
- 4.5.0.I20141029-2000 (count: 732)
- 4.6.0.I20150916-2000 (count: 669)
- 4.6.0.I20160405-0800 (count: 616)
- 4.5.0.I20141210-2000 (count: 575)
- 4.5.0.I20150203-1300 (count: 567)
- 4.5.2.M20160128-1800 (count: 521)
- 4.3.2.M20140221-1700 (count: 510)

Visualisation of the most used Eclipse Build IDs in the dataset:

Warning: Removed 2 rows containing missing values (position_stack).

Repartition of most used Eclipse Build IDs in dataset



Eclipse Product

- Description: The Eclipse product impacted by the exception.
- Type: Factors

The possible values found in the dataset for this attributes are:

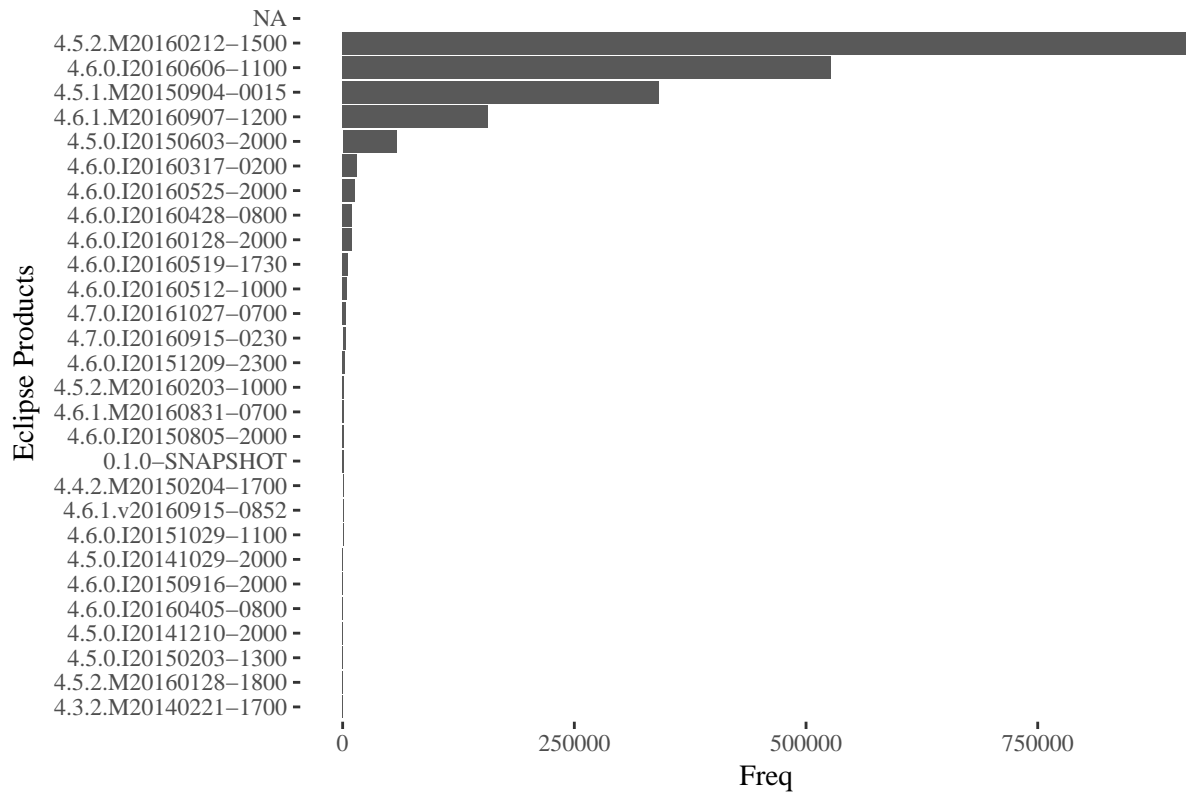
There are 150 different values found in the dataset for this attribute. The following table and bar plot only display the values with more than 500 occurrences:

- 4.5.2.M20160212-1500 (count: 914332)
- 4.6.0.I20160606-1100 (count: 527265)
- 4.5.1.M20150904-0015 (count: 341505)
- 4.6.1.M20160907-1200 (count: 156663)
- 4.5.0.I20150603-2000 (count: 58043)
- 4.6.0.I20160317-0200 (count: 15343)
- 4.6.0.I20160525-2000 (count: 13521)
- 4.6.0.I20160428-0800 (count: 10014)
- 4.6.0.I20160128-2000 (count: 9864)
- 4.6.0.I20160519-1730 (count: 5565)
- 4.6.0.I20160512-1000 (count: 4798)
- 4.7.0.I20161027-0700 (count: 3607)
- 4.7.0.I20160915-0230 (count: 3164)
- 4.6.0.I20151209-2300 (count: 2821)
- 4.5.2.M20160203-1000 (count: 1689)
- 4.6.1.M20160831-0700 (count: 1666)
- 4.6.0.I20150805-2000 (count: 1484)
- 0.1.0-SNAPSHOT (count: 1429)
- 4.4.2.M20150204-1700 (count: 999)
- 4.6.1.v20160915-0852 (count: 938)
- 4.6.0.I20151029-1100 (count: 904)
- 4.5.0.I20141029-2000 (count: 732)
- 4.6.0.I20150916-2000 (count: 669)
- 4.6.0.I20160405-0800 (count: 616)
- 4.5.0.I20141210-2000 (count: 575)
- 4.5.0.I20150203-1300 (count: 567)
- 4.5.2.M20160128-1800 (count: 521)
- 4.3.2.M20140221-1700 (count: 510)

Visualisation of the most used Eclipse Build IDs in the dataset:

```
## Warning: Removed 2 rows containing missing values (position_stack).
```

Repartition of most used Eclipse Products in dataset

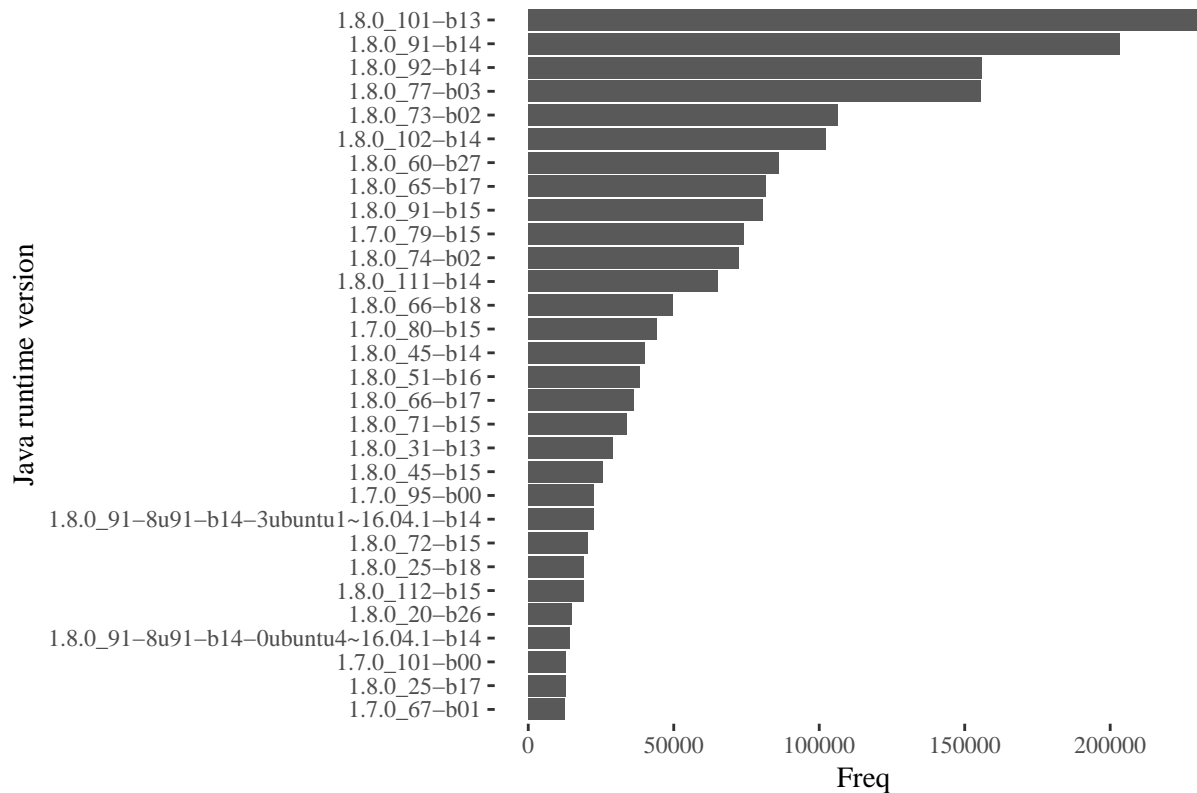


Java runtime version

- Description: The Java runtime of the host.
- Type: Factors

There are 709 different values found in the dataset for this attribute. The following bar plot only displays the values with more than 500 occurrences:

Repartition of top Java runtime versions in dataset



Comment Quality

- Description: An estimate of the user comment's quality (throughfulness). User comments help people better understand the context of the exception.
- Type: Factors

Distribution:

- HIGH 340
- MEDIUM 337
- LOW 255
- UNKNOWN 2083047

Using the dataset

Reading CSV file

Reading file from incidents_extract.csv.

```
myincidents <- read.csv(file.in, header=T)
myincidents[,c("bug", "status")] <- NULL
```

There are 20 columns and 2083979 entries in this dataset:

```
ncol(myincidents)
```

```
## [1] 20
```

```
nrow(myincidents)
```

```
## [1] 2083979
```

Names of columns:

```
names(myincidents)
```

```
## [1] "kind"           "severity"       "pluginId"
## [4] "pluginVersion" "code"          "statusFingerprint"
## [7] "message"       "statusSeverity" "fingerprint"
## [10] "fingerprint2"  "timestamp"     "savedOn"
## [13] "osgiArch"      "osgiOs"        "osgiOsVersion"
## [16] "osgiWs"        "eclipseBuildId" "eclipseProduct"
## [19] "javaRuntimeVersion" "commentQuality"
```

Using time series (xts)

The dataset needs to be converted to a xts object. We can use one of the 2 dates: `timestamp` or `savedOn`.

```
require(xts)
myp.xts <- xts(x = myincidents, order.by = parse_iso_8601(myincidents$savedOn))
```

Plot time series

Plot the number of weekly saves (attribute `savedOn`).

Weekly number of Incidents timestamp

