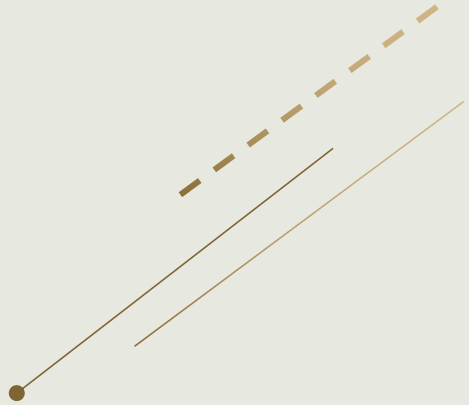
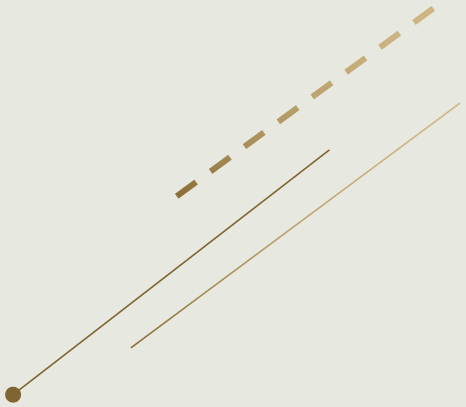


Web Usage Mining and User Preference

Bolong Zhang



AGENDA

- Introduction & Principle
- Process
- Result

Introduction & Principle

Web Usage Mining

Aim: Find User Browsing Behavior

Introduction & Principle

User - Web Relation ?

User -> IP Address

Browsing Behavior -> URL

Collaborative Filtering

To creat a IP - URL matrix

	URL 1	URL 2	URL 3	URL 4	URL 5	URL 6
IP 1	12	5		49	123	75
IP 2			2			
IP 3	23		64	54	125	54
IP 4	6		32			

IP X browse URL Y, then Matrix(X,Y) <= Matrix(X,Y)+1

Process

Raw Data Format:

<IP> <Time> <URL> <Status>

Raw Data: 5 IPs, 260 URLs

4508	10.129.2.1	[10/Nov/2017:15:55:47	GET /css/style.css HTTP/1.1	304
4509	10.131.2.1	[10/Nov/2017:15:55:47	GET /js/vendor/modernizr-2.8.3.min.js H	304
4510	10.130.2.1	[10/Nov/2017:15:55:47	GET /js/vendor/jquery-1.12.0.min.js HTT	304
4511	10.131.0.1	[10/Nov/2017:15:55:47	GET /bootstrap-3.3.7/js/bootstrap.min.j	304
4512	10.129.2.1	[10/Nov/2017:15:55:48	GET /fonts/fontawesome-webfont.woff2?v=	304
4513	10.129.2.1	[10/Nov/2017:15:55:51	POST /process.php HTTP/1.1	302
4514	10.129.2.1	[10/Nov/2017:15:55:52	GET /home.php HTTP/1.1	200
4515	10.129.2.1	[10/Nov/2017:15:55:53	GET /bootstrap-3.3.7/js/bootstrap.js HT	304
4516	10.129.2.1	[10/Nov/2017:15:55:53	GET /js/vendor/moment.min.js HTTP/1.1	304
4517	10.131.2.1	[10/Nov/2017:20:27:18	GET /robots.txt HTTP/1.1	404
4518	10.129.2.1	[11/Nov/2017:08:49:49	GET /robots.txt HTTP/1.1	404
4519	10.131.2.1	[11/Nov/2017:08:49:53	GET /sign.php HTTP/1.1	200
4520	10.129.2.1	[11/Nov/2017:10:33:18	GET /robots.txt HTTP/1.1	404
4521	10.131.0.1	[11/Nov/2017:12:19:05	GET / HTTP/1.1	302
4522	10.131.0.1	[11/Nov/2017:12:19:05	GET /login.php HTTP/1.1	200

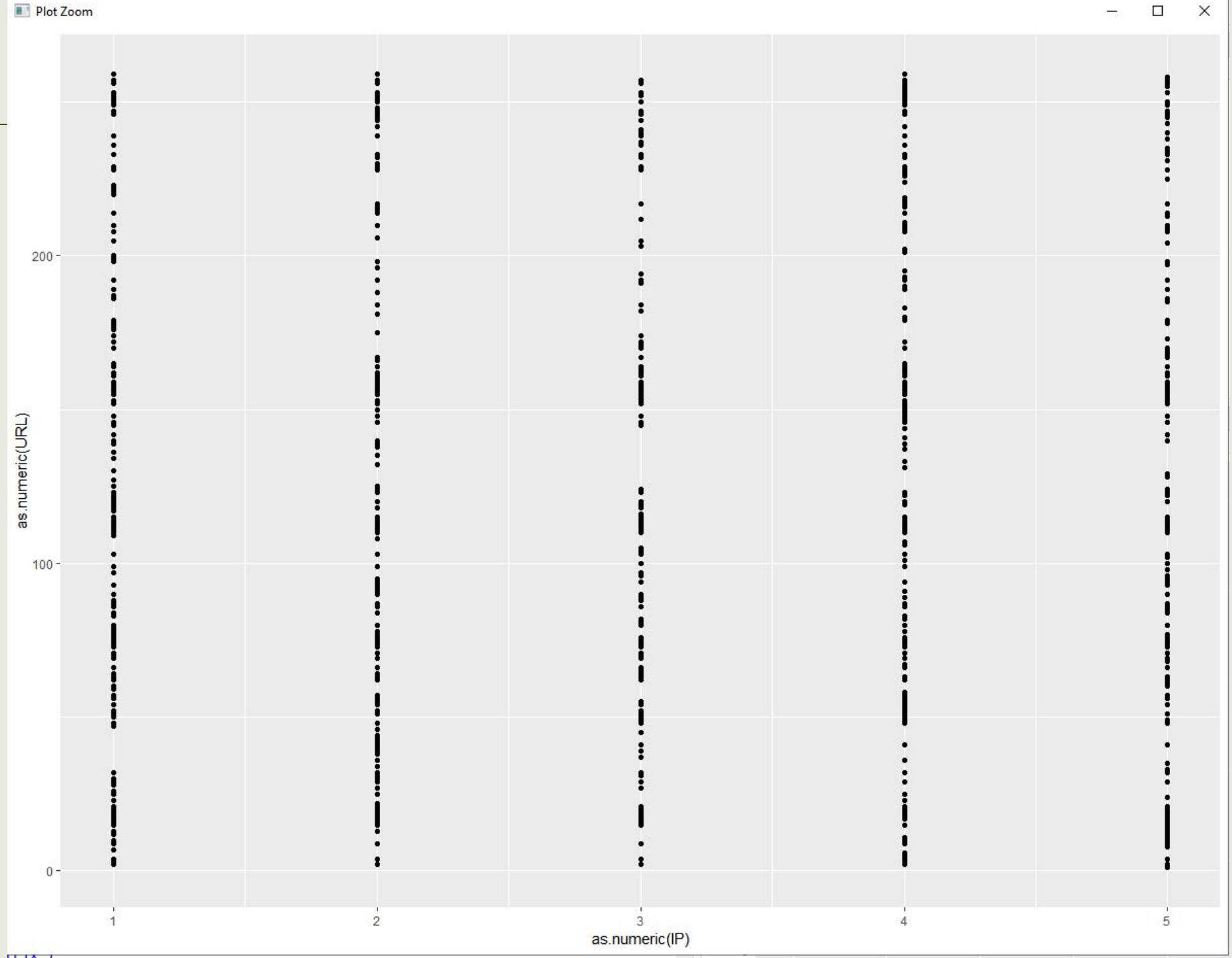
Origin Rows: 10787

After Cleaning: 10574

Process

X - User

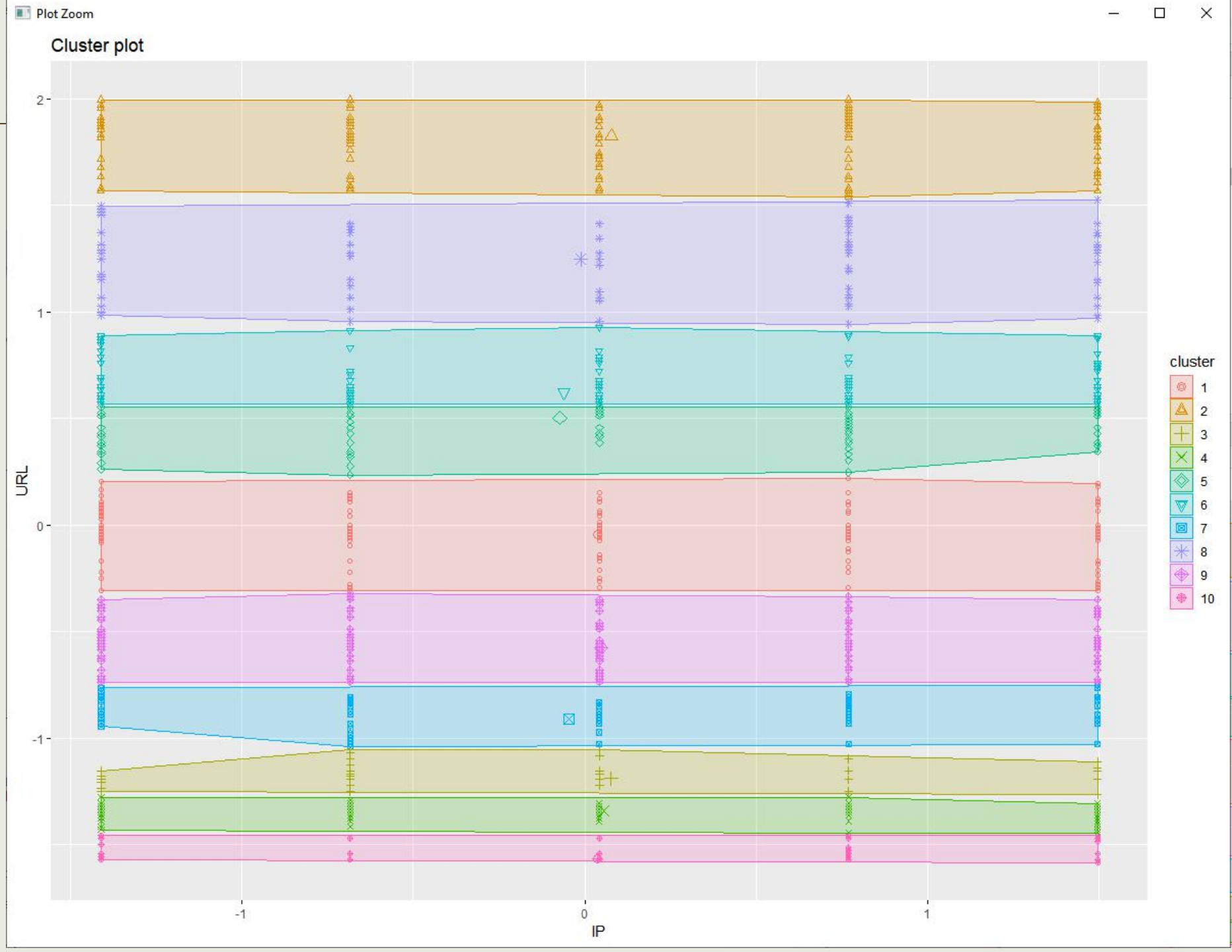
Y - Web



Process

X - User

Y - Web



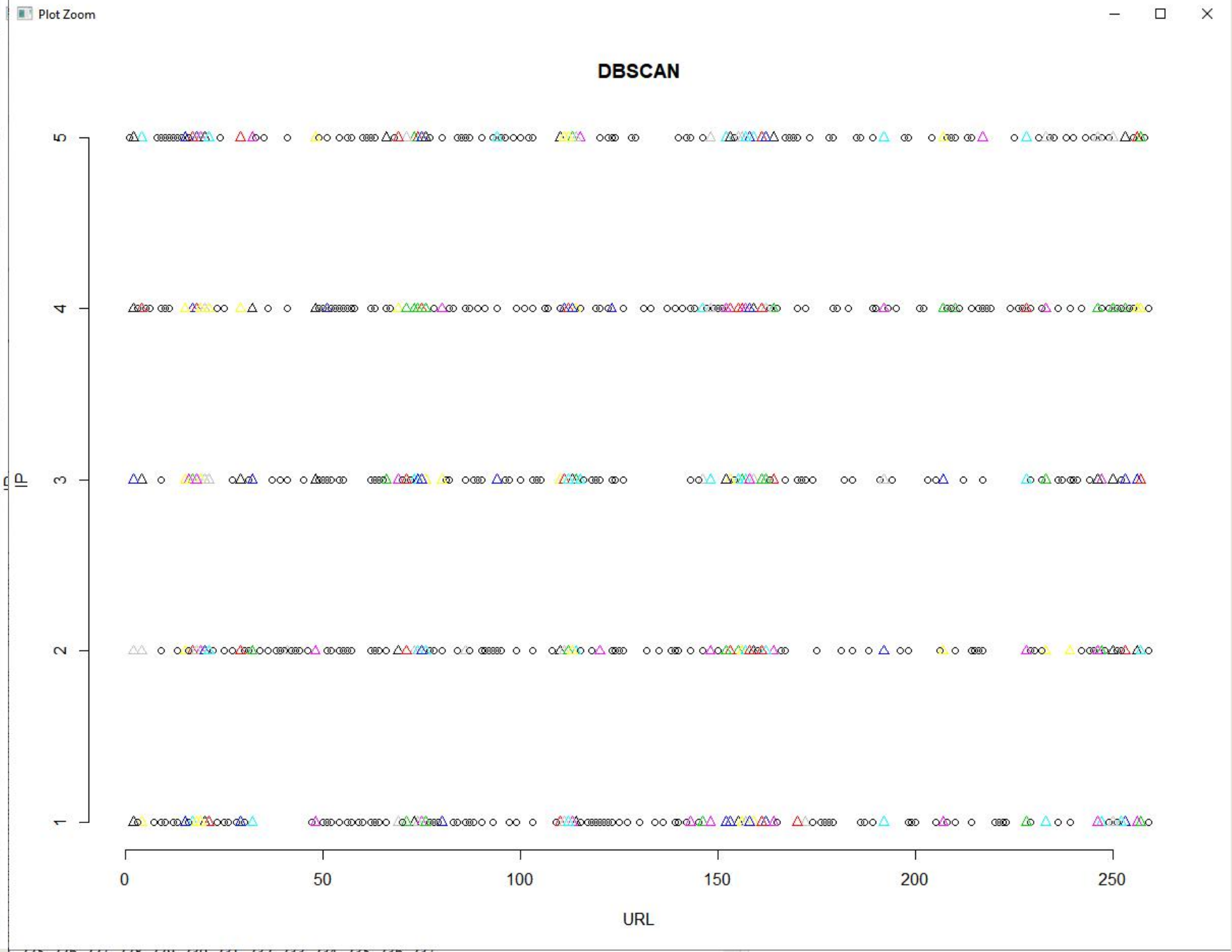
Process

X - Web

Y - URL

eps > 1

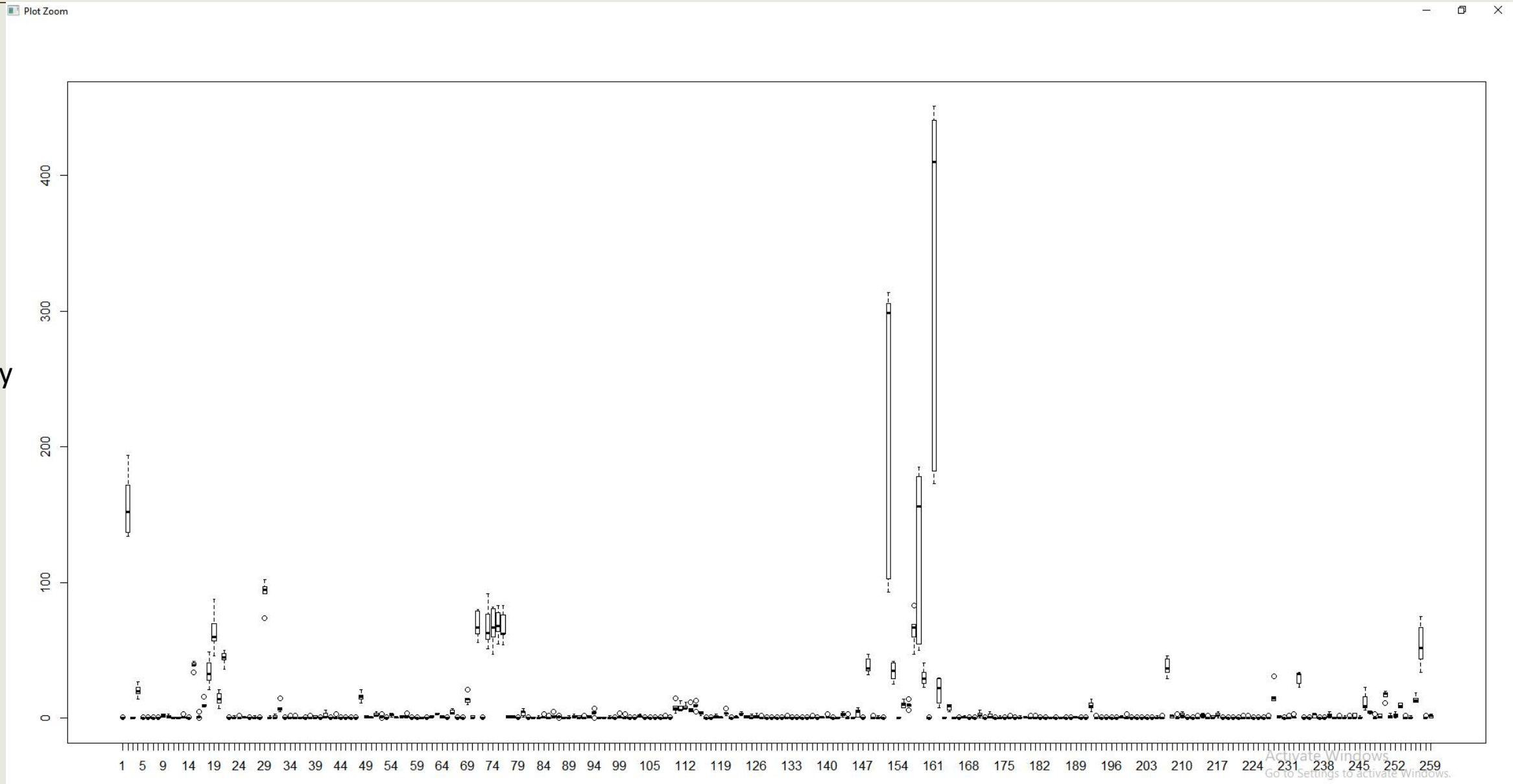
32 clusters



Process

X - Web

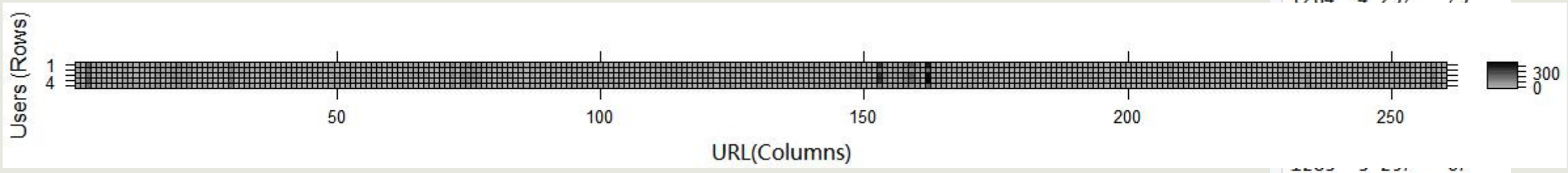
Y - Quantity



Process

X - Y -> (X, Y)

	IP	URL	Freq
804	4	161	451
803	3	161	441
801	1	161	410
758	3	152	314
756	1	152	306
759	4	152	299
9	4	2	194
786	1	158	185
802	2	161	182
788	3	158	178
805	5	161	173
6	1	2	172
789	4	158	156
8	3	2	152
10	5	2	137
7	2	2	134
757	2	152	103
145	5	29	102
144	4	29	97
143	3	29	95
760	5	152	93
141	1	29	92
364	4	73	92
94	4	19	88
373	3	75	83
379	4	76	83
783	3	157	83
366	1	74	82
369	4	74	81
353	3	71	80
351	1	71	79
374	4	75	78
363	3	73	77
378	3	76	76
1284	4	257	75
371	1	75	64



Result

User Based collaborative filtering

```
> as(pre,  "list" )
[[1]]
[1] "120" "66"  "80"

[[2]]
[1] "146" "66"  "80"

>
```

Item Based collaborative filtering

```
> as(pre,  "list" )
[[1]]
[1] "66" "80" "94"

[[2]]
[1] "66" "80" "94"

>
```

END

您的内容打在这里 或者通过复制您的文本