



## Customer segmentation – LifeCycle Grids with R

📅 16/02/2015 (<http://analyzecore.com/2015/02/16/customer-segmentation-lifecycle-grids-with-r/>)

👤 admin (<http://analyzecore.com/author/admin/>) 📁 LifeCycle Grids

(<http://analyzecore.com/category/lifecycle-grids/>), R language (<http://analyzecore.com/category/r-language/>), RFM (<http://analyzecore.com/category/rfm/>), Segmentation

(<http://analyzecore.com/category/segmentation/>), Visualization

(<http://analyzecore.com/category/visualization/>)

I want to share a very powerful approach for customer segmentation in this post. It is based on customer's lifecycle, specifically on **frequency** and **recency** of purchases. The idea of using these metrics comes from the **RFM analysis**. Recency and frequency are very important behavior metrics. We are interested in frequent and recent purchases, because frequency affects client's lifetime value and recency affects retention. Therefore, these metrics can help us to understand the current phase of the client's lifecycle. When we know each client's phase, we can split customer base into groups (segments) in order to:

- understand the state of affairs,
- effectively using marketing budget through accurate targeting,

- use different offers for every group,
- effectively using email marketing,
- increase customers' life-time and value, finally.

For this, we will use a matrix called **LifeCycle Grids**. We will study how to process initial data (transaction) to the matrix, how to visualize it, and how to do some in-depth analysis. We will do all these steps with the **R programming language**.

Let's create a data sample with the following code:

**click to expand R code**

The head of our data sample looks like:

	orderId	clientId	product	gender	orderdate
1	1	254	a	female	2012-04-03
2	1	254	b	female	2012-04-03
3	1	254	c	female	2012-04-03
4	1	254	b	female	2012-04-03
5	2	151	a	female	2012-01-31
6	2	151	b	female	2012-01-31

You can see that there is a gender of customer in the table. We will use it as an example of some in-depth analysis later. I recommend you to use any additional features, that you have, for seeking insights. It can be source of client, channel, campaign, geo data and so on.

A few words about **LifeCycle Grids**. It is a matrix with 2 dimensions:

- **frequency**, which is expressed in number of purchased items or placed orders,
- **recency**, which is expressed in days or months since the last purchase.

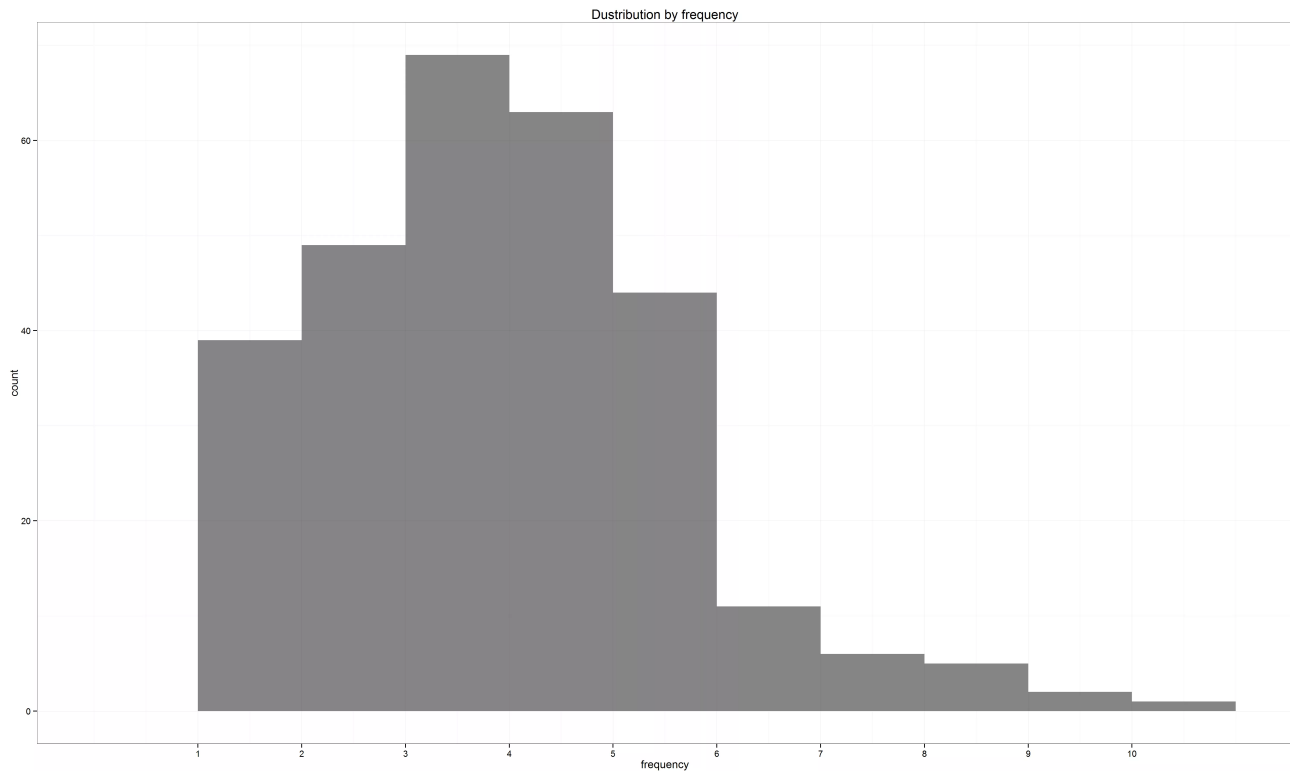
The first step is to think about suitable grids for your business. It is impossible to work with infinite segments. Therefore, we need to define some boundaries of frequency and recency, which should help us to split customers into homogeneous groups (segments). The analysis of the distribution of the frequency and the recency in our data set combined with the knowledge of business aspects can help us to find suitable boundaries.

Therefore, we need to calculate two values:

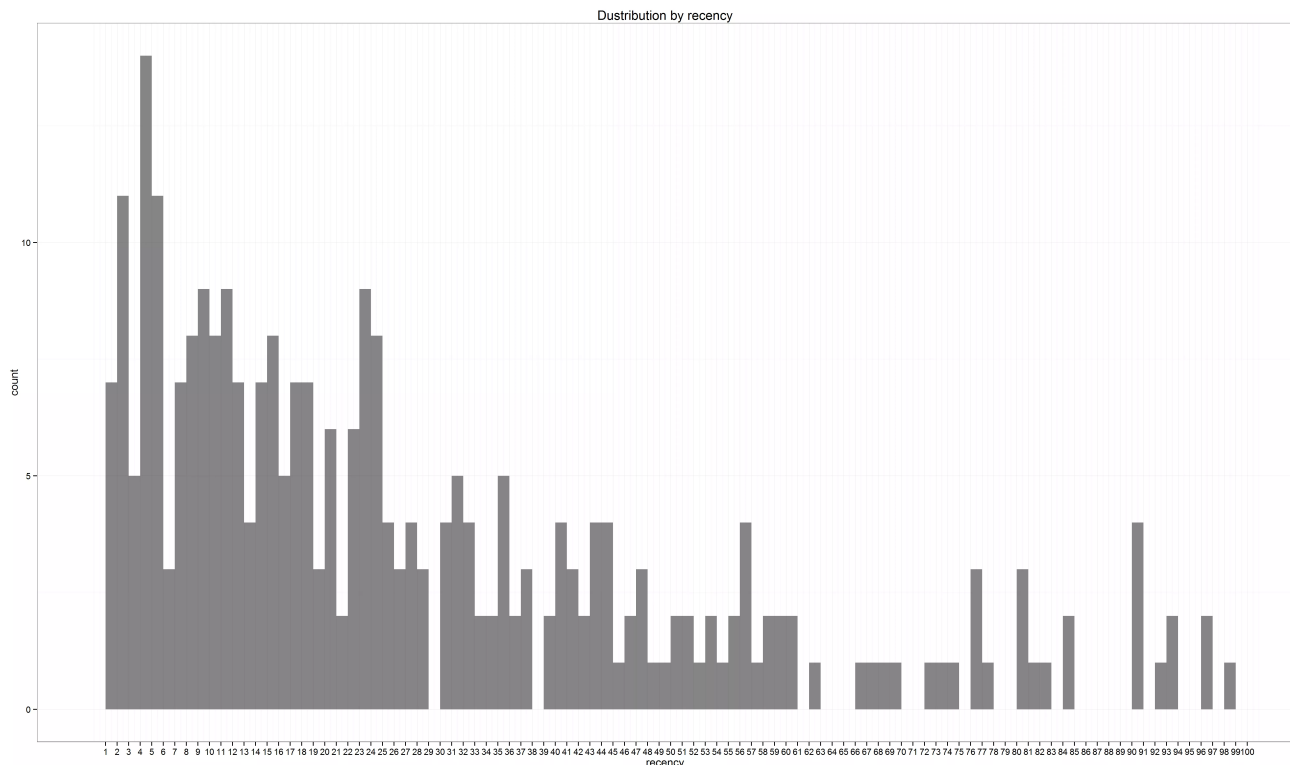
- number of orders that were placed by each client (or in some cases, it can be the number of items),
- time lapse from the last purchase to the reporting date.

Then, plot the distribution with the following code:

**click to expand R code**



([http://i0.wp.com/analyzecom/wp-content/uploads/2015/02/lcg\\_dist\\_1.png](http://i0.wp.com/analyzecom/wp-content/uploads/2015/02/lcg_dist_1.png))



([http://i2.wp.com/analyzecom/wp-content/uploads/2015/02/lcg\\_dist\\_2.png](http://i2.wp.com/analyzecom/wp-content/uploads/2015/02/lcg_dist_2.png))

Early behavior is most important, so finer detail is good there. Usually, there is a significant difference between customers who bought 1 time and those who bought 3 times, but is there any difference between customers who bought 50 times and other who bought 53 times? That is why it makes sense to set boundaries from lower values to higher gaps. We will use the following boundaries:

- for frequency: 1, 2, 3, 4, 5, >5,
- for recency: 0-6, 7-13, 14-19, 20-45, 46-80, >80

Next, we need to add segments to each client based on the boundaries. Also, we will create new variable 'cart', which includes products from the last cart, for doing in-depth analysis.

**click to expand R code**

We have everything need to create **LifeCycle Grids**. We need to combine clients into segments with the following code:

**click to expand R code**

The classic matrix can be created with the following code:

**click to expand R code**

	segm.freq	>80 days	46-80 days	20-45 days	14-19 days	7-13 days	0-6 days
1	>5	0	1	6	4	7	7
2	5	0	1	15	8	9	11
3	4	0	1	21	9	20	12
4	3	0	10	25	11	10	13
5	2	2	17	18	3	4	5
6	1	12	13	7	2	2	3

([http://i2.wp.com/analyzecore.com/wp-content/uploads/2015/02/lcg\\_matrix.png](http://i2.wp.com/analyzecore.com/wp-content/uploads/2015/02/lcg_matrix.png))

However, I suppose a good visualization is obtained through the following code:

**click to expand R code**



([http://i0.wp.com/analyzecore.com/wp-content/uploads/2015/02/lcg\\_1\\_1.png](http://i0.wp.com/analyzecore.com/wp-content/uploads/2015/02/lcg_1_1.png))

I've added colored borders for a better understanding of how to work with this matrix. We have four quadrants:

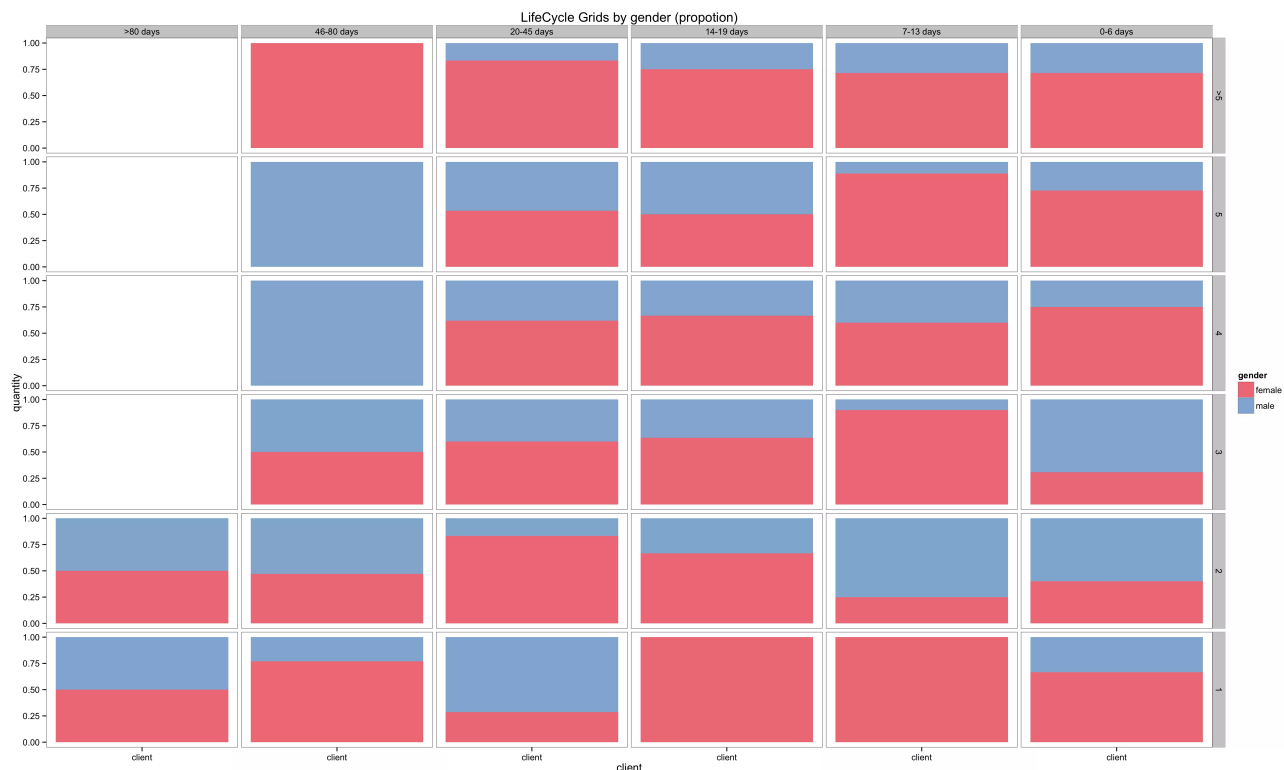
- **yellow** – here are our **best customers**, who have placed quite a few orders and made their last purchase recently. They have higher value and higher potential to buy again. We have to take care of them.
- **green** – here are our **new clients**, who placed several orders (1-3) recently. Although they have lower value, they have potential to move into the yellow zone. Therefore, we have to help them move into the right quadrant (yellow).
- **red** – here are our **former best customers**. We need to understand why they are former and, maybe, try to reactivate them.
- **blue** – here are our **onetime-buyers**.

Does it make sense to make the same offer to all of these customers? Certainly, it doesn't! It makes sense to create different approaches not only for each quadrant, but for border cells as well.

What I really like about this model of segmentation is that it is stable and alive simultaneously. It is alive in terms of customers flow. Every day, with or without purchases, it will provide customers flow from one cell to another. And it is stable in terms of working with segments. It allows to work with customers who are on the same lifecycle phase. That means you can create suitable campaigns / offers / emails for each or several close cells and use them constantly.

Ok, it's time to study how we can do some in-depth analysis. R allows us to create subsegments and visualize them effectively. It can be helpful to distribute each cell via some features. For instance, there can distribute customers by gender. For the other example, where our products have different lifecycles, it can be helpful to analyze which product/s was/were in the last cart or we can combine these features. Let's do this with the following code:

**click to expand R code**



([http://i1.wp.com/analyzecore.com/wp-content/uploads/2015/02/lcg\\_21.png](http://i1.wp.com/analyzecore.com/wp-content/uploads/2015/02/lcg_21.png)) or even:

**click to expand R code**



([http://i0.wp.com/analyzecore.com/wp-content/uploads/2015/02/lcg\\_31.png](http://i0.wp.com/analyzecore.com/wp-content/uploads/2015/02/lcg_31.png))

([http://analyzecore.com/wp-content/uploads/2015/02/lcg\\_3.png](http://analyzecore.com/wp-content/uploads/2015/02/lcg_3.png))

Therefore, there is a lot of space for creativity. If you want to know much more about LifeCycle Grids and strategies for working with quadrants, I highly recommend that you read **Jim Novo's** works, e.g. this blogpost (<http://blog.jimnovo.com/2007/04/25/engagement-customers/>).

Thank you for reading this!

#### Share this:

**in** (<http://analyzecore.com/2015/02/16/customer-segmentation-lifecycle-grids-with-r/?share=linkedin&nb=1>)

146

**✈** (<http://analyzecore.com/2015/02/16/customer-segmentation-lifecycle-grids-with-r/?share=twitter&nb=1>)

**f** (<http://analyzecore.com/2015/02/16/customer-segmentation-lifecycle-grids-with-r/?share=facebook&nb=1>)

17

**G+** (<http://analyzecore.com/2015/02/16/customer-segmentation-lifecycle-grids-with-r/?share=google-plus-1&nb=1>)

**👤** (<http://analyzecore.com/2015/02/16/customer-segmentation-lifecycle-grids-with-r/?share=reddit&nb=1>)

**p** (<http://analyzecore.com/2015/02/16/customer-segmentation-lifecycle-grids-with-r/?share=pinterest&nb=1>)

1

**✉** (<http://analyzecore.com/2015/02/16/customer-segmentation-lifecycle-grids-with-r/?share=email&nb=1>)

#### Like this:



4 bloggers like this.

#### Related



(<http://analyzecore.com/2015/04/01/cohort-analysis-and-lifecycle-grids-mixed-segmentation-with-r/>)  
Cohort Analysis and LifeCycle Grids mixed



(<http://analyzecore.com/2015/02/19/customer-segmentation-lifecycle-grids-clv-and-cac-with-r/>)  
Customer segmentation - LifeCycle Grids, CLV and CAC



(<http://analyzecore.com/2015/08/27/measuring-business-health-with-delta-lifecycle-grids-and-r/>)  
Measuring business health with Delta LifeCycle Grids

segmentation with R  
(<http://analyzecore.com/2015/04/01/cohort-analysis-and-lifecycle-grids-mixed-segmentation-with-r/>)  
In "Cohort analysis"

with R  
(<http://analyzecore.com/2015/02/19/customer-segmentation-lifecycle-grids-clv-and-cac-with-r/>)  
In "Customer life-time value"

and R  
(<http://analyzecore.com/2015/08/27/measuring-business-health-with-delta-lifecycle-grids-and-r/>)  
In "LifeCycle Grids"

LifeCycle Grids (<http://analyzecore.com/tag/lifecycle-grids/>) R (<http://analyzecore.com/tag/r/>)

RFM (<http://analyzecore.com/tag/rfm/>) Segmentation (<http://analyzecore.com/tag/segmentation/>)

Visualization (<http://analyzecore.com/tag/visualization/>)

17 Comments

AnalyzeCore

Login ▾

Recommend 2

Share

Sort by Best ▾



Join the discussion...



**Jason Aizkalns** • 5 months ago

Great stuff. Some food for thought, if you add two additional factors/labels for recency/frequency, you can get away with not adding the manual color boundaries. For example:

```
lcg$segm.rec.label <- ifelse(lcg$segm.rec %in% c(">80 days", "46-80 days", "20-45 days"))
lcg$segm.freq.label <- ifelse(lcg$segm.freq %in% c(">5", "5", "4"), "frequent", "infrequent")

# Take a look at the this...
# interaction(lcg$segm.rec.label, lcg$segm.freq.label)

ggplot(lcg, aes(x=client, y=quantity, fill=interaction(segm.rec.label, segm.freq.label)))
  theme_bw() +
  theme(panel.grid = element_blank())+
  geom_bar(stat='identity', alpha=0.6) +
  geom_text(aes(y=max(quantity)/2, label=quantity), size=4) +
  facet_grid(segm.freq ~ segm.rec) +
  ggtitle("LifeCycle Grids")
```

This generates the following graph:







^ | v • Reply • Share ›



**AnalyzeCore** Mod → Jason Aizkalns • 5 months ago

Jason, thanks for the comment! It is really useful!

^ | v • Reply • Share ›



**Jason Aizkalns** → AnalyzeCore • 5 months ago

No problem - great stuff -- keep it up. Another alternative would be to use `geom\_rect` and fill the background see the following

<http://stackoverflow.com/q/984...>

You can probably use that approach and still color by other factors (such as Gender and Product in your examples).

Just getting back into this "space" -- you may want to look at (if you haven't already) the BTYD package:

<https://cran.r-project.org/web...>

<https://cran.r-project.org/web...>

<http://www.slideshare.net/matt...>

^ | v • Reply • Share ›



**AnalyzeCore** 24 Jason Aizkalns 17 6 146 1

Sure! Will think how to use these idea! Thanks!

^ | v • Reply • Share ›



**Scott Horvath** • 7 months ago

Thanks for the post. How do you know whether the 'best customers' are 'best customer' who has bought five items?

^ | v • Reply • Share ›



**AnalyzeCore** Mod → Scott Horvath • 7 months ago

Thank you for the question, Scott! There are several perspectives on the surface:

1) Using orders/purchases instead of items. You can use orders instead of items as in my example. In this case, it doesn't matter how many items were bought, but the fact of purchase matters. The more purchases the better client. Furthermore, customers could place several orders in dozen minutes in online business. Therefore, if you think this was the one contact with your company you can calculate more than 1 order per day (e.g. 3 orders) as the one purchase.

2) Using items with suitable boundaries. If a customer purchases several items

frequently, you need to define boundaries in suitable for your business way. Then

(http://anal...)  
CUSTOMER LIFETIME  
SEQUENCE OF  
VALUE  
Sequence of  
shopping carts in-  
depth analysis with  
LifeCycle Grids,  
CLV...

re.co re.co

m/20 m/20

15/01 15/02

128/s 119/c

eque usto

Customer segmentation - LifeCycle Grids with R - AnalyzeCore - data are beautiful, data are a story frequently you need to define boundaries in suitable for your business way. They mer- could be 1-100 items, 101-1000 and so on. Actually, LifeCycle Grids approach starts with analyzing customer's behavior and identifying recency and frequency carts- segm boundaries.

3) Combining with Cohort analysis. This is the most accurate way I think. You can combine LCG with Cohort analysis (by first purchase date). It means you would see both the best and the new customers who did 5 purchases in the same cell depth on different cohorts. Read more about in my following post <http://analyzecore.com/2015/04...> - lifecy

Hope this helps.

^ | v • Reply • Share ›



**Scott Horvath** → AnalyzeCore • 7 months ago

Many thanks for your response.

I just want to confirm that my interpretation of the chart [0] is correct. with- clv-

I read that 4 unique customers (best customers) made more than five orders and- with their last order occurring 14-19 days ago.

I also read that 11 unique customers (new customers) made 3 orders with their last order occurring 14-19 days ago. If we consider these 11 customers, event cac- that have been flagged as 'new', how do you know that the first (or second) of s/) with- their three orders wasn't made 46-80 days ago? If they had, then I believe r/) they would be incorrectly flagged as 'new'.

Very eager to hear your thoughts!

Scott

[0]



[see more](#)

^ | v • Reply • Share ›



**AnalyzeCore** Mod → Scott Horvath • 7 months ago

You are right. The brand new customers are only in the cell 1 purchase with their last order occurring 0-6 days ago. The colored blocks are indicative and were used for explanation of the approach and you can use another definition for cells or for blocks. For instance, if you read the post I mentioned you would find another example:

- new customer (1-2 purchases and 0-60 days recency),
- under risk new customer (1-2 purchases and 61-180 days recency),

- 1x buyer (1-2 purchases and >180 days recency),
  - engaged customer (3-4 purchases and 0-60 days recency),
  - under risk engaged customer (3-4 purchases and 61-180 days recency),
  - former engaged customer (3-4 purchases and >180 days recency),
  - best customer (>4 purchases and 0-60 days recency),
  - under risk best customer (>4 purchases and 61-180 days recency),
  - former best customer (>4 purchases and >180 days recency).
- Therefore, it is up to you how to combine cells and name them.

But I want to pay your attention that LCG approach is about lifecycle phase that we try to identify through purchases and their recency. That is why the customer, who made 3 orders with the last order occurring 14-19 days ago can be 'new' for you even the first order was 80 days ago, because in our relationships were 3 orders only and this wasn't enough to say that customer is loyal and we can't offer something special for. This model is simple and flexible but needs some work for adapting to exact business model.

^ | v • Reply • Share ›



**Scott Horvath** → AnalyzeCore • 7 months ago

Really appreciate the response! I think I will go and try and make something like this. I will reference this post, too!

^ | v • Reply • Share ›



**AnalyzeCore** Mod → Scott Horvath • 7 months ago

Thank you, Scott!

^ | v • Reply • Share ›



**max** • a year ago

Hi and thank you for your post , i have used the same example like you but i get the graph without colored borders?

why?

thanks

^ | v • Reply • Share ›



**AnalyzeCore** Mod → max • a year ago

I've added colored borders manually in order to give a description how to read the grids. Thank you!

^ | v • Reply • Share ›



**max** → AnalyzeCore • a year ago

ok thanks , but How to know for example if one client change from group to

 |  • Reply • Share ›**AnalyzeCore** Mod  max • a year ago

There are several ways. You can find some ideas in this post

<http://analyzecore.com/2015/04...>

Further, I'm going to publish a separate post about.

 |  • Reply • Share ›**max**  AnalyzeCore • a year ago

thank you , i'll try to understand

Great post! My only suggestions would be:

to work with a real data in order to show the importance of your method

Thanks

 |  • Reply • Share ›**AnalyzeCore** Mod  max • a year ago

Thank you for the advice! But real data are confidential in most of the cases.

 |  • Reply • Share ›**max**  AnalyzeCore • a year ago

Hi,

for your next post,

How can I add variable CAC and grossmargin?

1.CAC:cost acquisition customer : for instance you send mail for your customer so the cost is always zero

2.grossmargin:that means how much I earn per product

I'm very grateful if you can explain to me How can I get these variables knowing that I work with a real dataset

Thanks!

 |  • Reply • Share ›

## ALSO ON ANALYZECORE

**Customer segmentation – LifeCycle Grids, CLV and CAC with R**

5 comments • a year ago

**Cohort analysis: Retention Rate Visualization with R**

3 comments • 5 months ago



**AnalyzeCore** — Sergey, thank for the good point! For me, CLV is a value of customer for whole lifetime. Therefore, if we speak



**ethen** — Thanks for the quick response, all is clear now!

◀ Sequence of shopping carts in-depth analysis with R – Sequence of events  
(<http://analyzecore.com/2015/01/28/sequence-carts-in-depth-analysis-with-r-events/>)

Customer segmentation – LifeCycle Grids, CLV and CAC with R ▶ (<http://analyzecore.com/2015/02/19/customer-segmentation-lifecycle-grids-clv-and-cac-with-r/>)



## FEEL FREE TO CONTACT ME VIA LINKEDIN



(<https://www.linkedin.com/in/sergeybryl/>)

**Sergey Bryl'** (<https://www.linkedin.com/in/sergeybryl/>)

**Data Scientist at Wargaming.net | Data are beautiful, data are a story**

bryl.serg@gmail.com <http://analyzecore.com/> Data are beautiful, data are a story for me. I enjoy working with data: creating models, predicting and finding insights. My real goal is to address [...]

## FOLLOW ME ON TWITTER

## Tweets by @AnalyzeCore



**AnalyzeCore**  
@AnalyzeCore

#Cohortanalysis: #Retention Rate #Visualization with #R  
[wp.me/p4bGVE-p0](http://wp.me/p4bGVE-p0)

**Cohort analysis: Retention Rate Visualization with R - AnalyzeCore - data are beauti...**  
When conducting Cohort Analysis, one of the most important measures is Customer Ret...  
[analyzecore.com](http://analyzecore.com)

10 Dec

[Embed](#)

[View on Twitter](#)

## SUBSCRIBE

Enter your email address to subscribe to this blog and receive notifications of new posts by email.

SUBSCRIBE

## CATEGORIES

📁 [Clustering \(http://analyzecore.com/category/clustering/\)](http://analyzecore.com/category/clustering/) (1)

📁 [Cohort analysis \(http://analyzecore.com/category/cohort-analysis/\)](http://analyzecore.com/category/cohort-analysis/) (6)

📁 [Customer life-time value \(http://analyzecore.com/category/customer-life-time-value/\)](http://analyzecore.com/category/customer-life-time-value/) (3)

📁 [Forecasting \(http://analyzecore.com/category/forecasting/\)](http://analyzecore.com/category/forecasting/) (1)

📁 [LifeCycle Grids \(http://analyzecore.com/category/lifecycle-grids/\)](http://analyzecore.com/category/lifecycle-grids/) (4)

📁 [R language \(http://analyzecore.com/category/r-language/\)](http://analyzecore.com/category/r-language/) (18)

📁 [Retention rate \(http://analyzecore.com/category/retention-rate/\)](http://analyzecore.com/category/retention-rate/) (1)

📁 RFM (<http://analyzecore.com/category/rfm/>) (2)

📁 Sales Funnel (<http://analyzecore.com/category/sales-funnel/>) (1)

📁 Segmentation (<http://analyzecore.com/category/segmentation/>) (4)

📁 Sentiment Analysis (<http://analyzecore.com/category/sentiment-analysis/>) (1)

📁 Sequential analysis (<http://analyzecore.com/category/sequential-analysis/>) (5)

📁 Twitter analysis (<http://analyzecore.com/category/twitter-analysis/>) (2)

📁 Visualization (<http://analyzecore.com/category/visualization/>) (16)

## USEFUL LINKS

R-bloggers  (<http://www.r-bloggers.com/>)

Marketing Productivity Blog by Jim Novo  (<http://blog.jimnovo.com/>)

## RSS

RSS - Posts (<http://analyzecore.com/feed/>)

---

sparkling Theme by Colorlib (<http://colorlib.com/>) Powered by WordPress (<http://wordpress.org/>)