About

The contents of this chapter are separate from the clustering guide

Contents

1 About

2 Terms

3 Sources

4 Contact

About

Credits

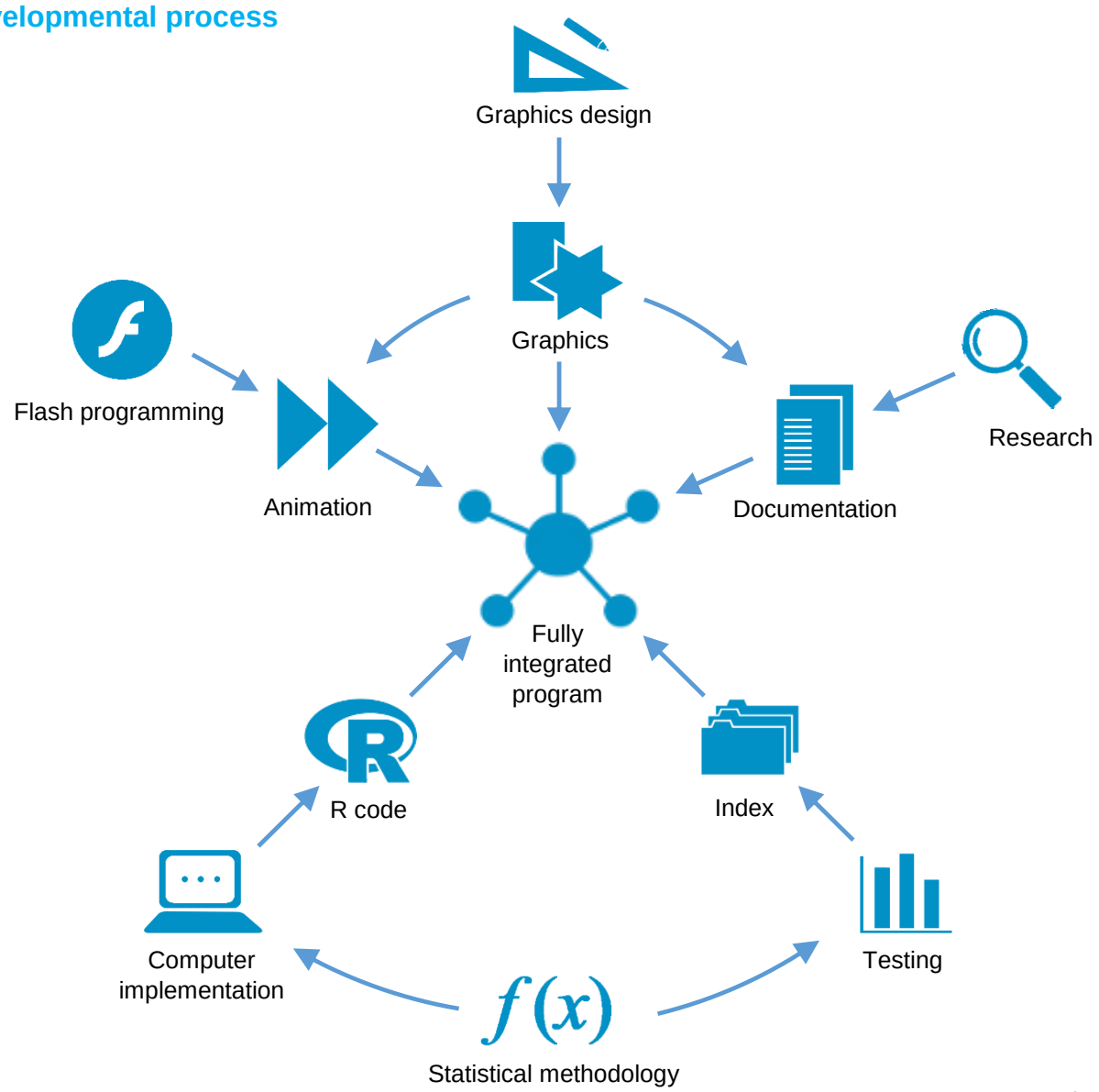
This guide was created by Yang Li for AbacusBio Ltd. in February 2016

All writing, graphics, programming, animation and overall design were done by Yang Li. Programs used include:

- Microsoft Word for writing
- GIMP for graphics
- Adobe Flash for animation
- R for statistical programming

Special thanks to Bram Visser and Peter Fennessy (supervisors)

Developmental process



Terms

Terms of use

When using this guide, keep the following in mind:

1. The creator (Yang Li) allows AbacusBio Ltd. to use, distribute and modify the content of the guide in absolutely any way

2. The creator/AbacusBio expressly disclaim any and all liabilities contingent or otherwise that may arise from the use of the information or recommendations in the guide or associated code. Without limiting the scope of this statement, the following apply:

    2.1 The creator/AbacusBio do not guarantee that the guide is 100% accurate. However, every effort has been made to ensure the accuracy of the guide, with every statement supported by sources (see 'Sources' section).

    2.2 The creator/AbacusBio are not responsible for any damage to or loss of data that may arise from misuse. However, none of the code provided has any chance of damaging data if used as directed.

    2.3 The creator/AbacusBio are not responsible if the user suffers any form of loss from interpreting data with an unsuitable method. *None* of the methods provided will work in all situations.

3. Bug reports and additional questions can be sent to the creator (see 'Contact' section), though the creator has no obligation or intention of maintaining/updating this program in the long term.

Sources

Sources of information

The following sources were used when researching for this guide. Some sources were used extensively, while others contributed less than a sentence (No sources were directly copied from). They are listed in the order that they were retrieved. Note that webpages may change.

- http://www.statsoft.com/Textbook/Cluster-Analysis
- http://stats.stackexchange.com/questions/133656/how-to-understand-the-drawbacks-of-k-means
- http://stackoverflow.com/questions/15376075/cluster-analysis-in-r-determine-the-optimal-number-of-clusters
- http://www.machinelearning.org/proceedings/icml2007/papers/216.pdf
- http://webcache.googleusercontent.com/search?q=cache:http://bvmengineering.ac.in/docs/published%2520papers/cpit/cpit/201405.pdf&gws_rd=cr&ei=H3JKVsL5A8fTmwWUz6eACA
- https://sites.google.com/site/vegclassmethods/statistical-analyses/classification-methods#TOC-Other-numerical-approaches
- https://books.google.co.nz/books?id=RQHB4_p3bJoC&pg=PA412&lpg=PA412
- https://en.wikipedia.org/wiki/DBSCAN
- https://en.wikipedia.org/wiki/K-medoids#Algorithms
- http://www.ijcta.com/documents/volumes/vol3issue5/ijcta2012030521.pdf
- http://zdb.ru.lv/conferences/3/VTR8_II_70.pdf
- http://slideplayer.com/slide/3200567/
- http://ijcai.org/papers13/Papers/IJCAI13-249.pdf
- http://www.tqmp.org/RegularArticles/vol09-1/p015/p015.pdf
- http://www.ijircce.com/upload/2014/january/7_A%20comparative.pdf
- http://www.ijraset.com/fileserve.php?FID=467
- https://cran.r-project.org/web/packages/vegclust/vignettes/VegetationClassification.pdf
- http://www.public.iastate.edu/~maitra/stat501/lectures/ModelBasedClustering.pdf
- http://www.slideshare.net/petitegeek/expectation-maximization-and-gaussian-mixture-models
- https://cran.r-project.org/web/packages/kernlab/vignettes/kernlab.pdf
- http://wheatoncollege.edu/lexomics/files/2012/08/How-to-Read-a-Dendrogram-Web-Ready.pdf
- http://www.unesco.org/webworld/idams/advguide/Chapt7_1_5.htm
- http://people.stat.sc.edu/Hitchcock/chapter6_R_examples.txt
- http://stackoverflow.com/questions/26019584/understanding-concept-of-gaussian-mixture-models?rq=1
- http://hameddaily.blogspot.co.nz/2015/03/when-not-to-use-gaussian-mixtures-model.html
- https://www.diva-portal.org/smash/get/diva2:198852/FULLTEXT01.pdf
- http://trace.tennessee.edu/cgi/viewcontent.cgi?article=2096&context=utk_gradthes
- http://blog.surveymethods.com/when-to-use-single-imputation-or-multiple-imputation/
- https://en.wikipedia.org/wiki/Imputation_(statistics)#Single_imputation
- https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2012/35_Japan.pdf
- http://machinelearningmastery.com/feature-selection-with-the-caret-r-package/
- http://www.r-bloggers.com/counting-clusters/
- https://en.wikibooks.org/wiki/Data_Mining_Algorithms_In_R/Clustering/Density-Based_Clustering
- http://www.osl.iu.edu/~chemuell/projects/presentations/optics-v1.pdf
- http://www.slideshare.net/rpiryani/optics-ordering-points-to-identify-the-clustering-structure

Contact

Contacting the creator

The creator can be contacted at the following email address:

combee.yang@yahoo.com

While the creator will make every effort to respond to questions and fix problems, note that:

1. The creator should only be contacted if the user has tried to fix the problem themselves
2. A response may be delayed depending on how busy the creator is
3. The creator may not be able to fix all problems

Additional help with code

General questions about R code can be sent to Bram Visser:

bvisser@abacusbio.co.nz