



Exon Junction Read Arches: Visualizing Data in R

Jennifer Chang

Who Am I?



Jennifer Chang

2nd Year PhD Student at Iowa State University
B.A. Computer Science & Biochemistry
(Cornell College)



Major Adviser: Dr Hui-Hsien Chou

Complex Computational Laboratory



Learning the way of life —
computationally.

Our mission is to facilitate and expediate research
in molecular biology with an integrated approach
combining computation sciences with life sciences.

:: our mission



Road Map (Outline)

- **Introduction**
 - Grammer of Graphics
 - Junction Reads
- **Goals & Developing “Arches”**
- **Current State & Future Plans**
 - Bioconductor & ggbio
- **Acknowledgements**



Introduction

How I got involved in Data Visualization

Grammer of Graphics

Exon Junction Reads

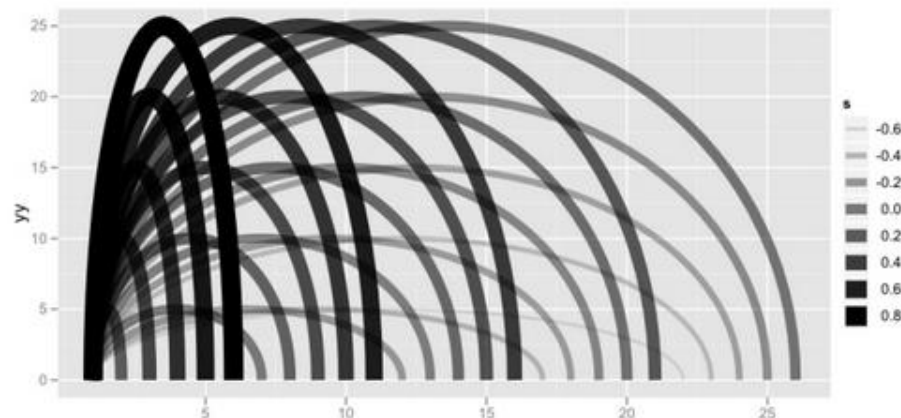


[Home](#) [Biological Background](#) [Current Code](#) [Statement of Goals](#)

White Breaks (overplotting?)

Posted on November 16, 2011

To create the arches, I originally was plotting 1000 points. (well 1002 to be exact.) I've reduced the number of points and the breaks in the arches have changed. I reduced the number of points to plot the arch to 50 (well 52 to be exact.) and the breaks seem to have disappeared.



Recent Posts

- [White Breaks \(overplotting?\)](#)
- [Use Alpha instead of Colour](#)
- [Example Showing Over Arches and Under Arches](#)
- [Adjustment to Line Width and Color Based on Jump Distance](#)
- [From ggplot to geom](#)

Archives

- [November 2011](#)
- [October 2011](#)

Categories

- [Uncategorized](#)

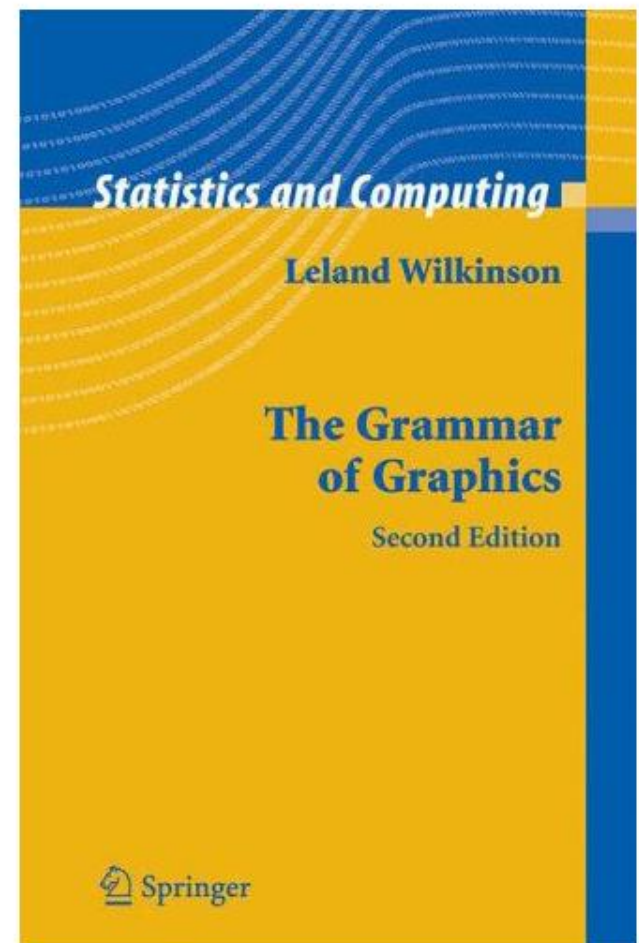
Meta

- [Register](#)
- [Log in](#)
- [Entries RSS](#)
- [Comments RSS](#)
- [WordPress.com](#)

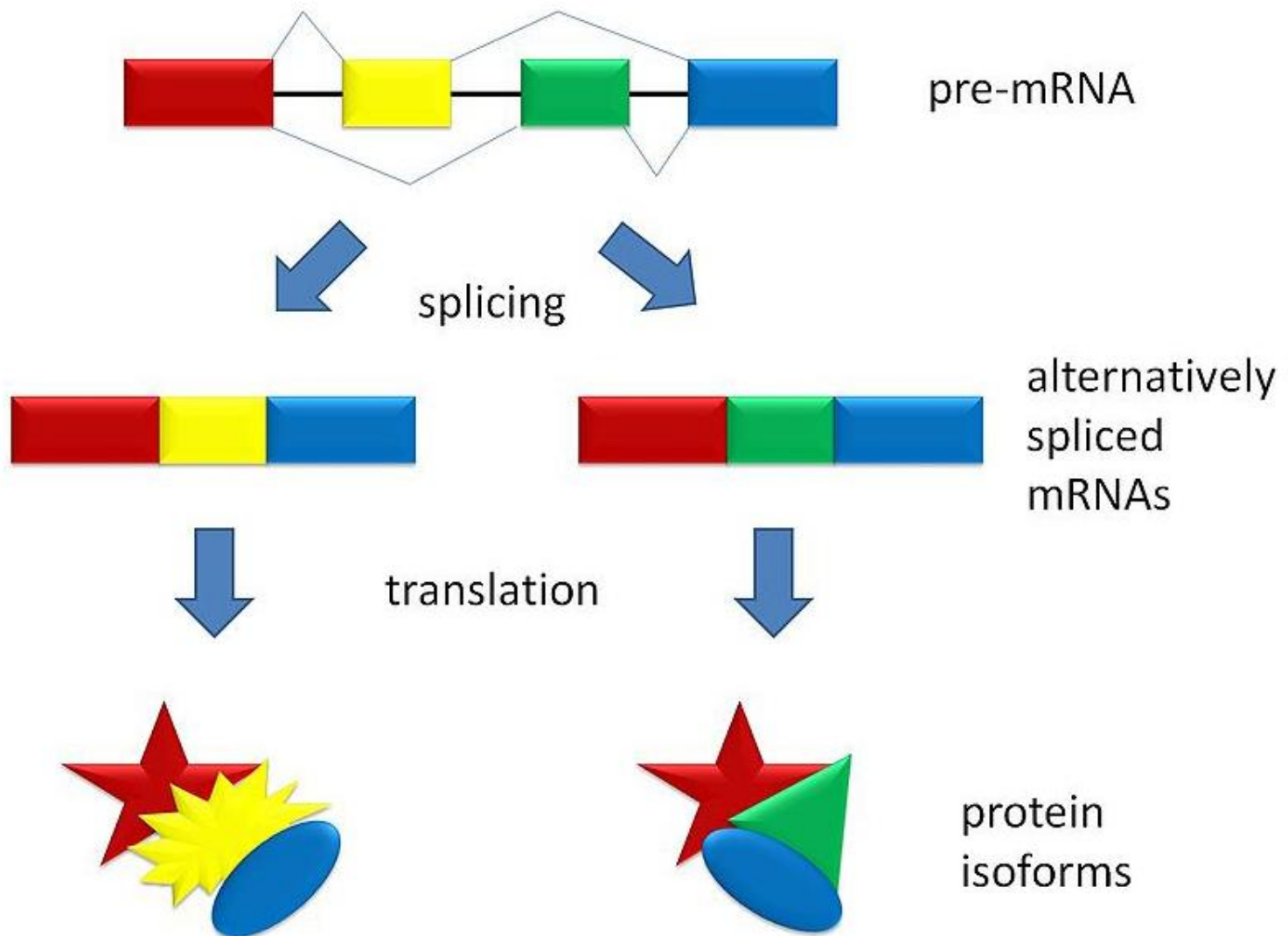
Grammar of Graphics

“grammar” for good
graphics

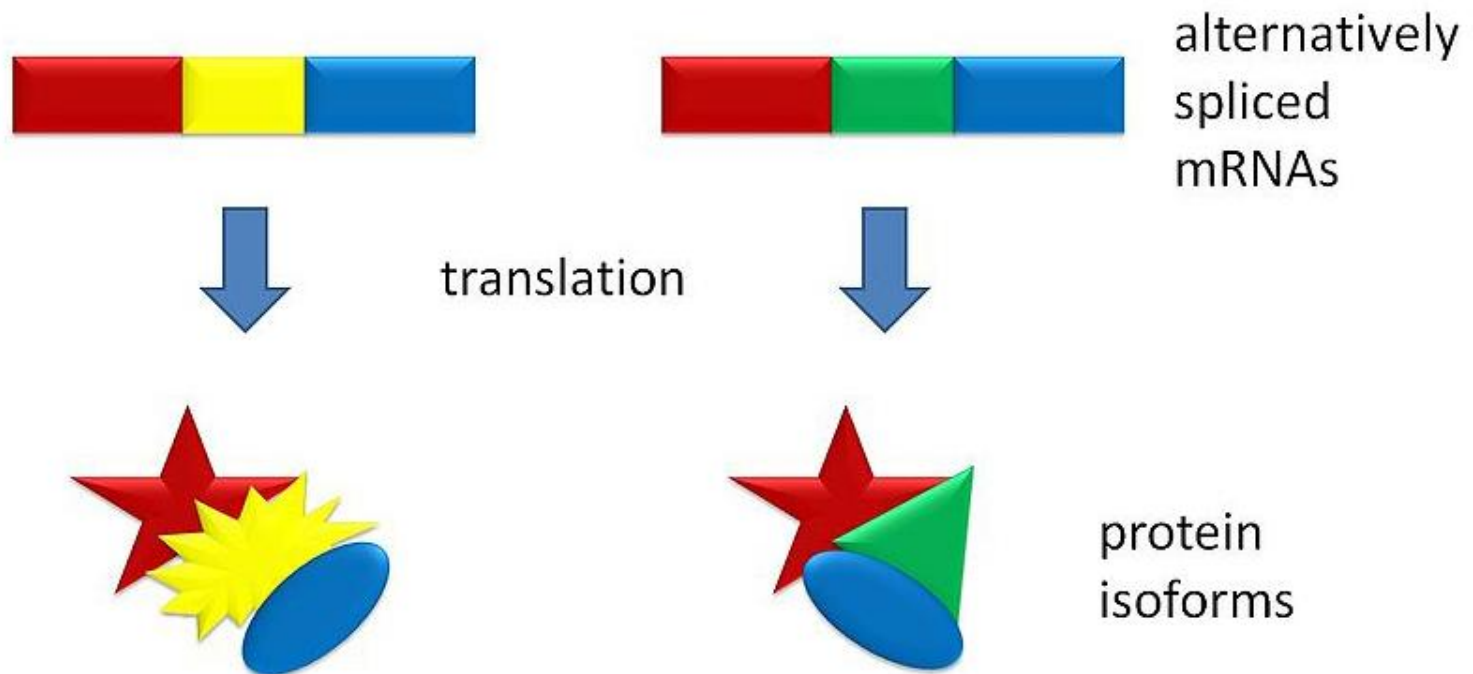
ggplot2 - Hadley Wickam



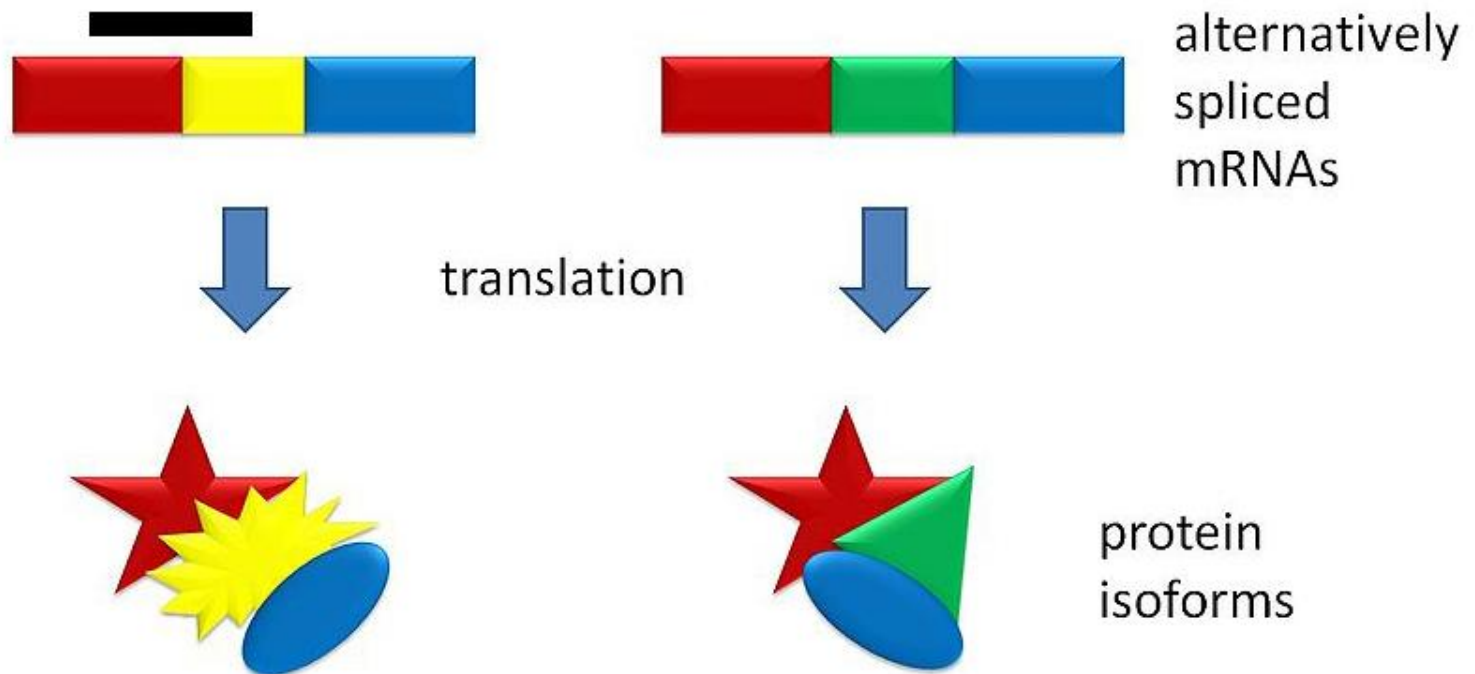
Exon Junction Reads



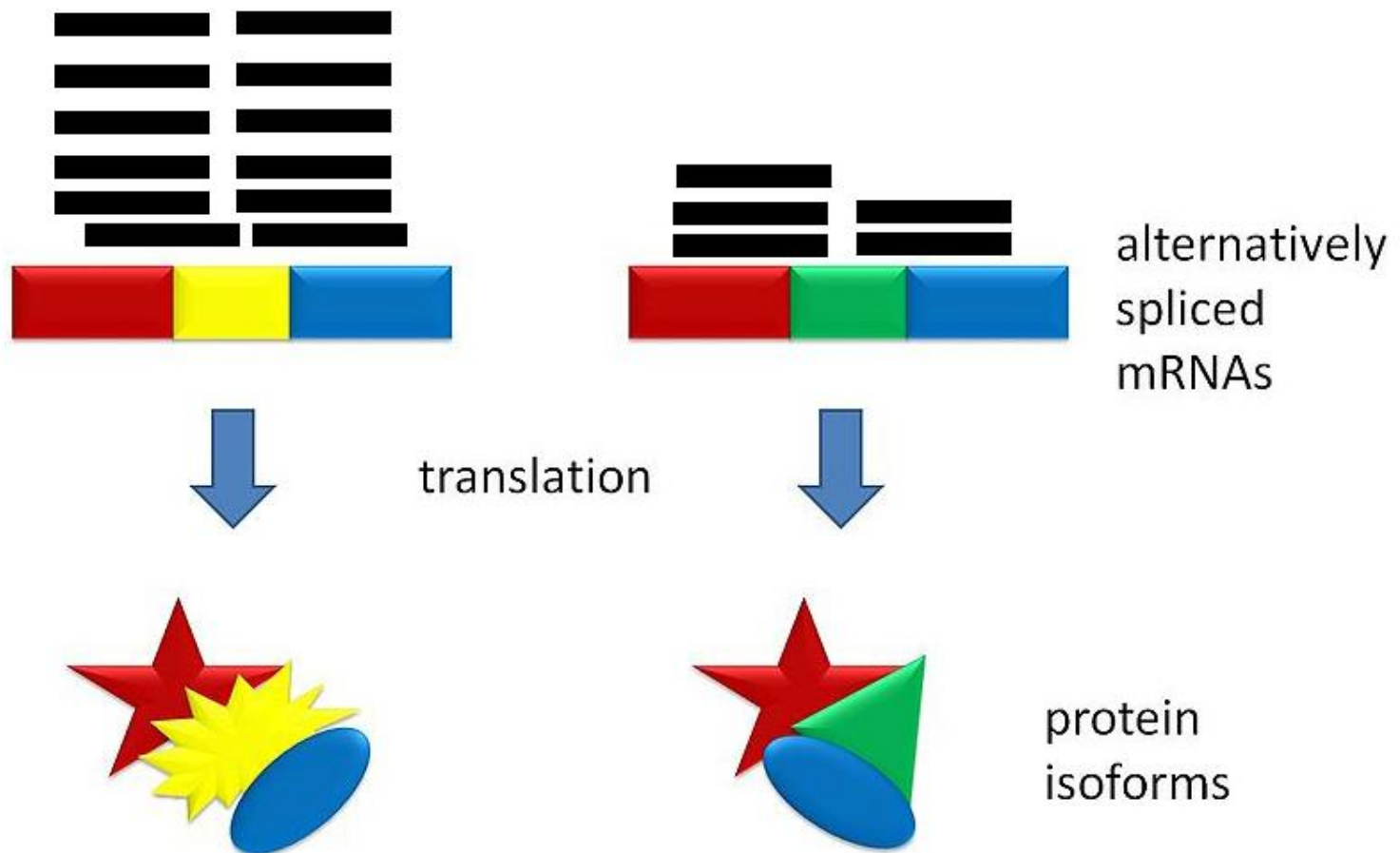
Exon Junction Reads



Exon Junction Reads



Exon Junction Reads

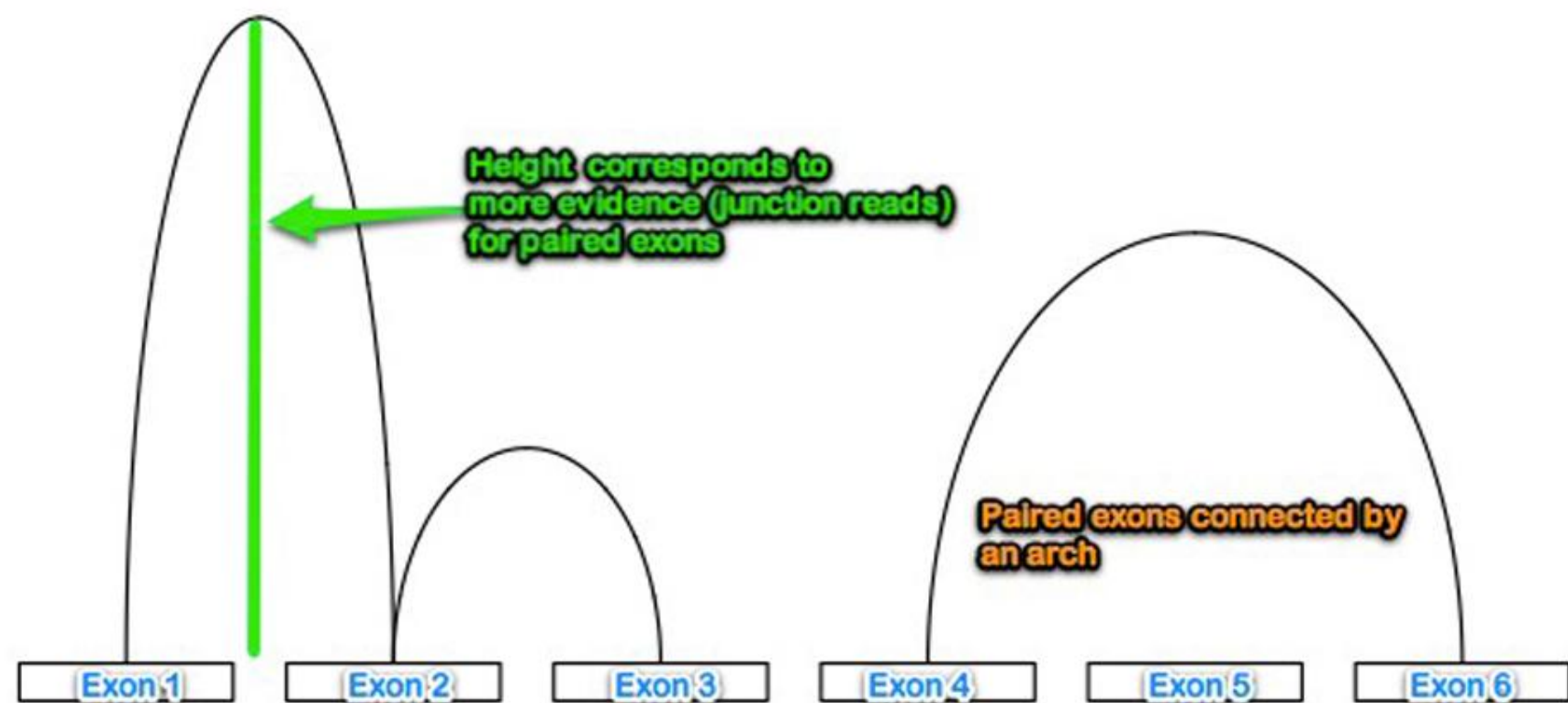


Goals:



- **Compare Gene Expression**
- **Better Understand Alternative Splicing**
- **Insights into Disease and Treatment**
- **Compare Different Treatment Groups**
- **Identify Correlation**
- **Visualize Alternative Splicing**

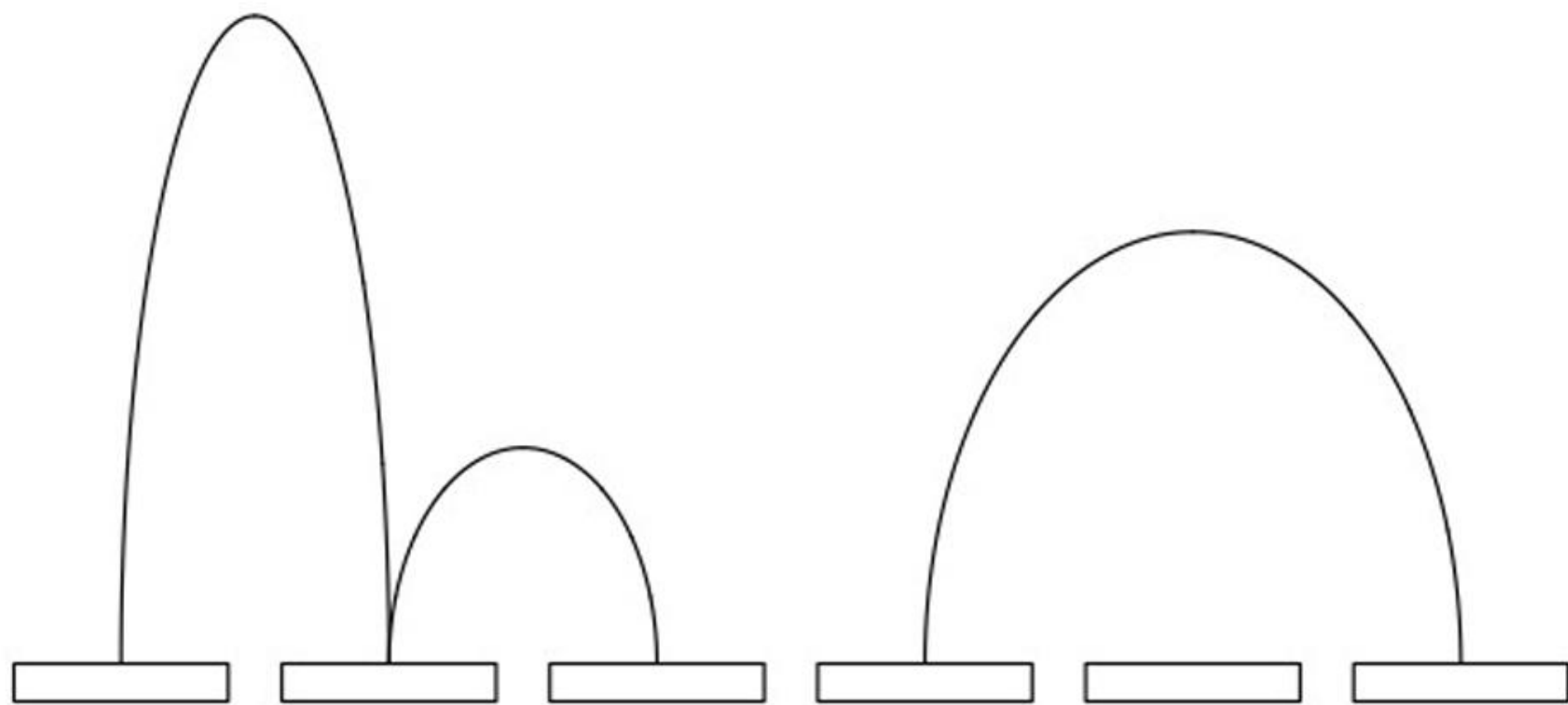
Goals:



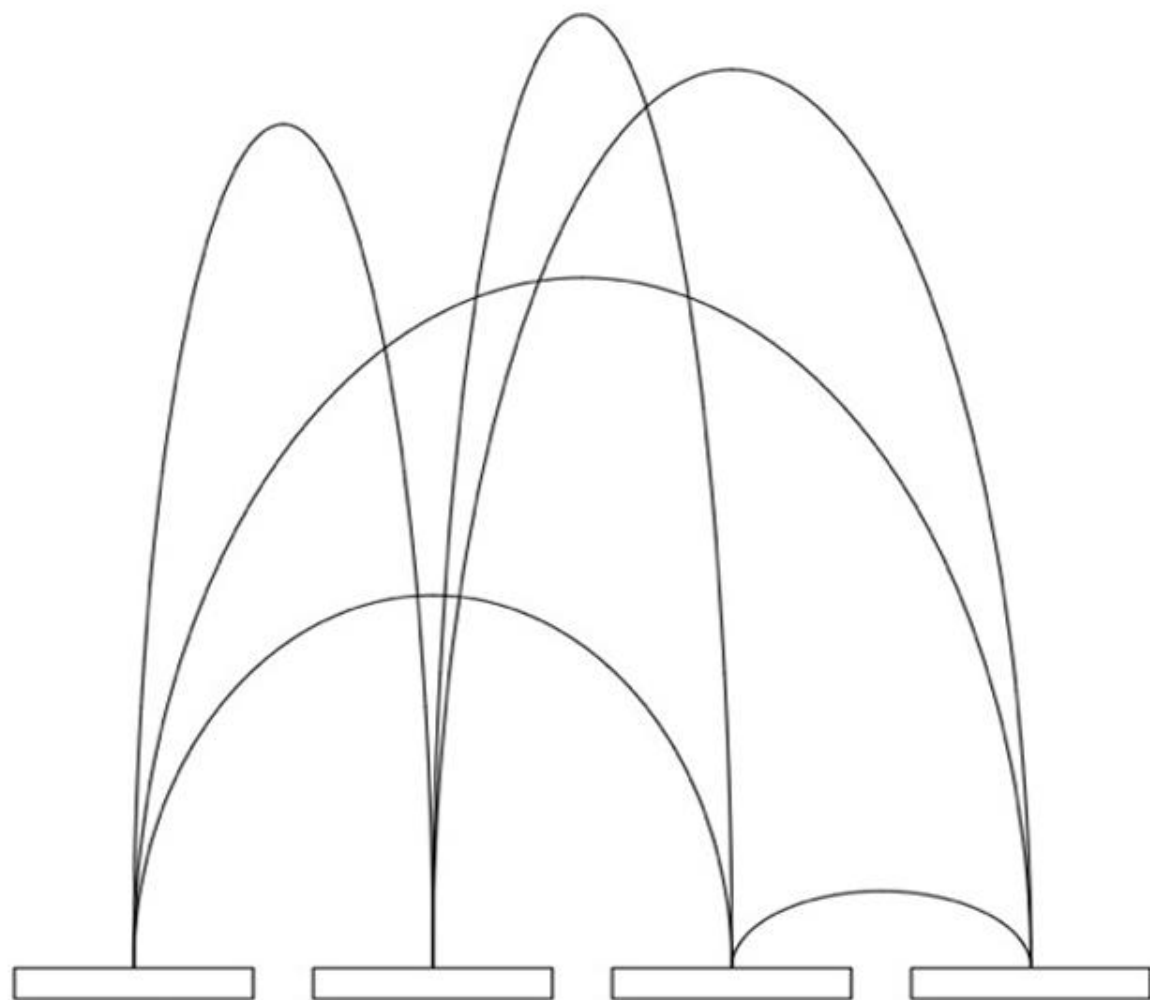
Implement it!!!

(just keep coding, just keep coding . . .)

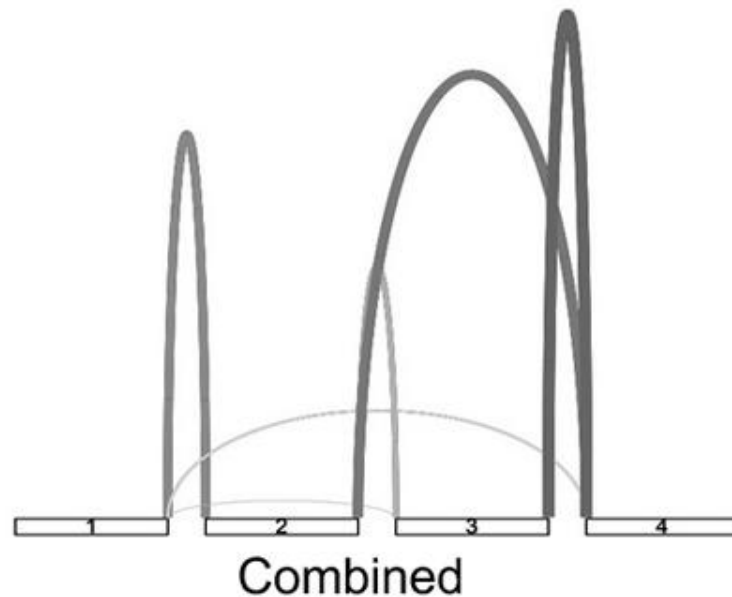
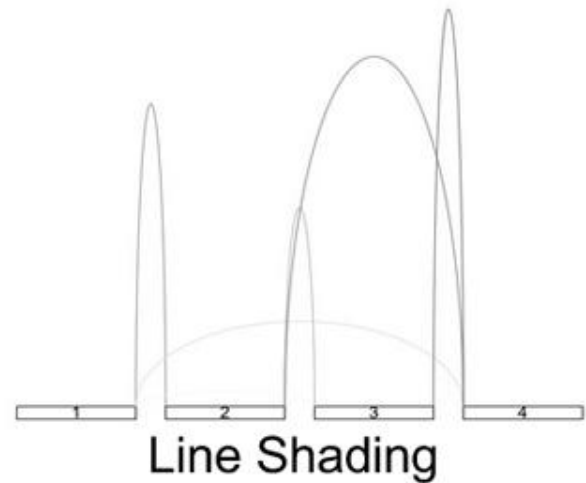
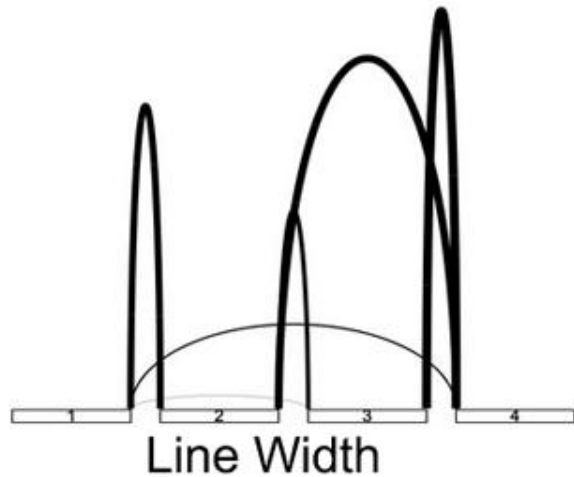
Development:



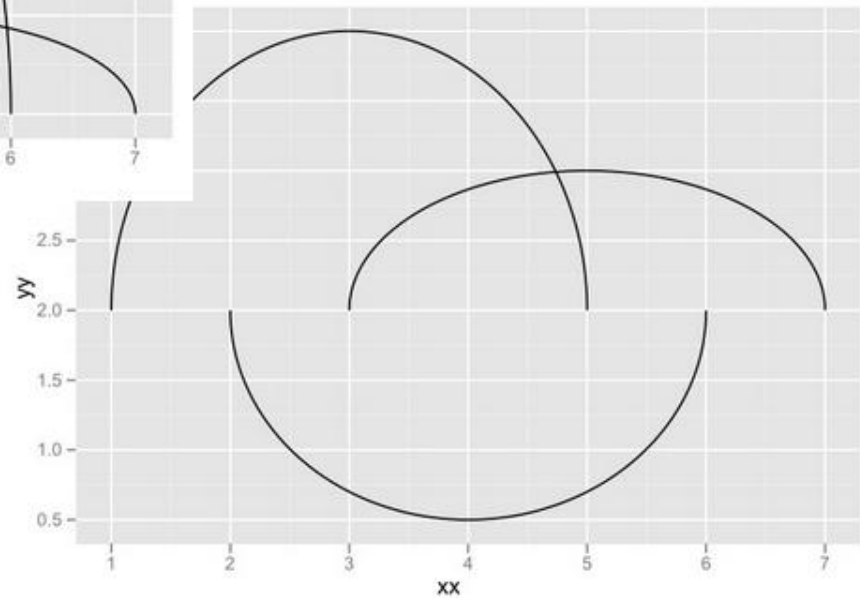
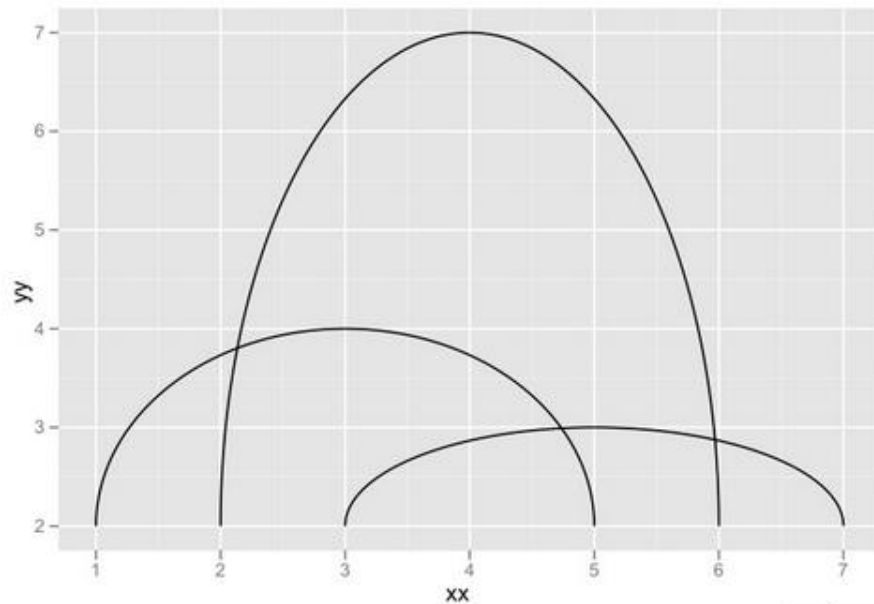
Development:



Development:



Development:



Development:

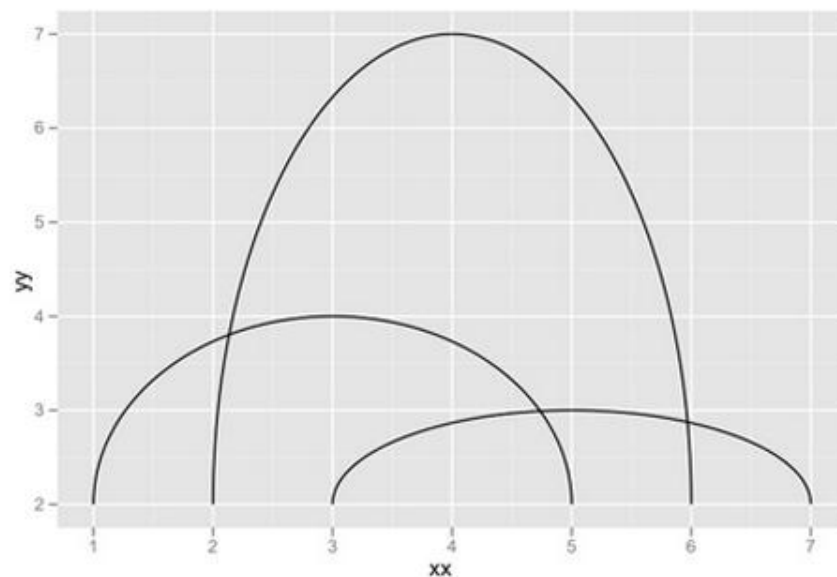
Changes in the Input

Take in end points instead of the midpoint

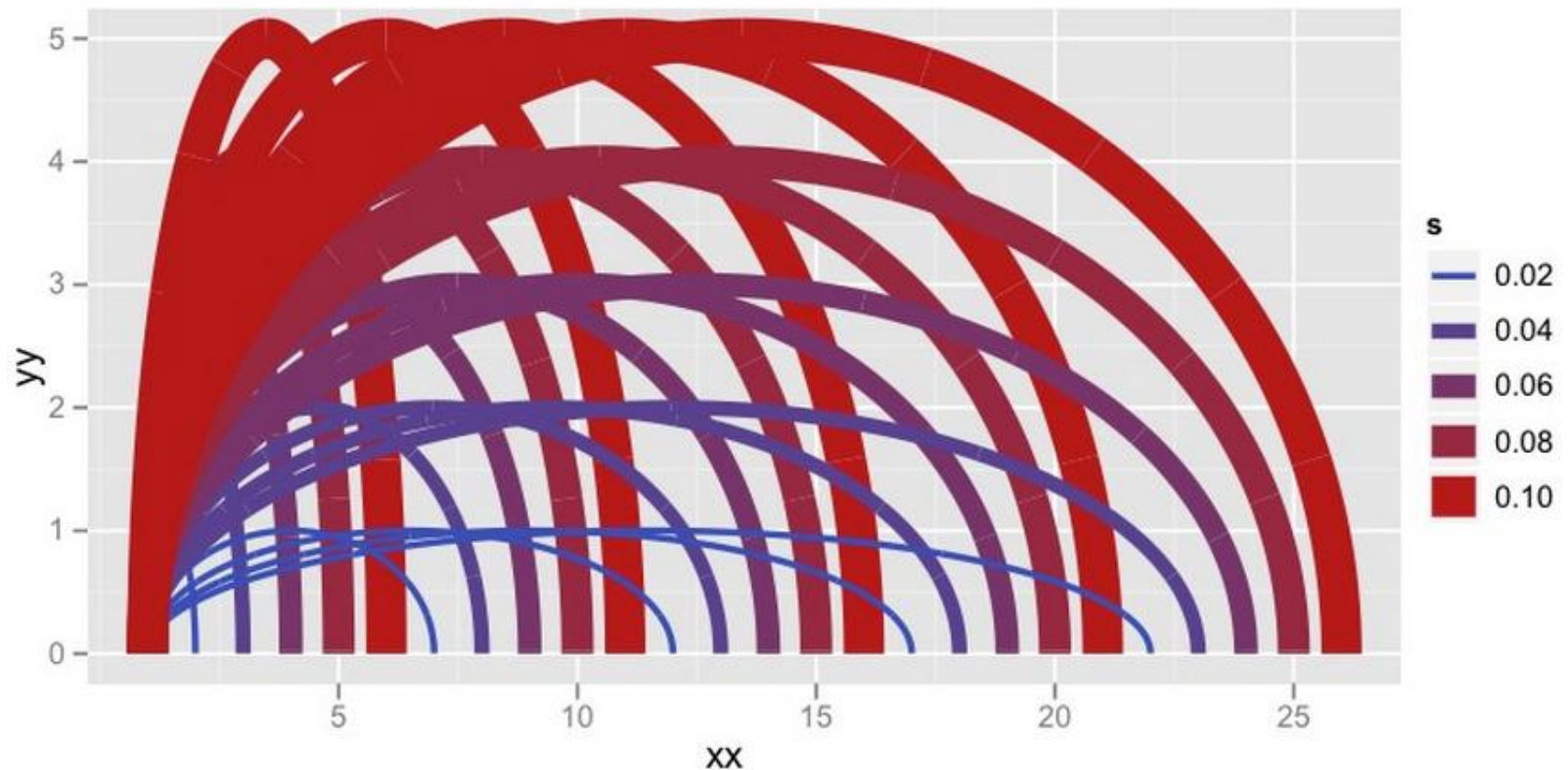
Map aesthetics and extended geom_line

Format of the Command:

```
ggplot(exons, junctionReads)+  
  geom_arch(data = , aes(x = , xend = , y =, height = ,))
```



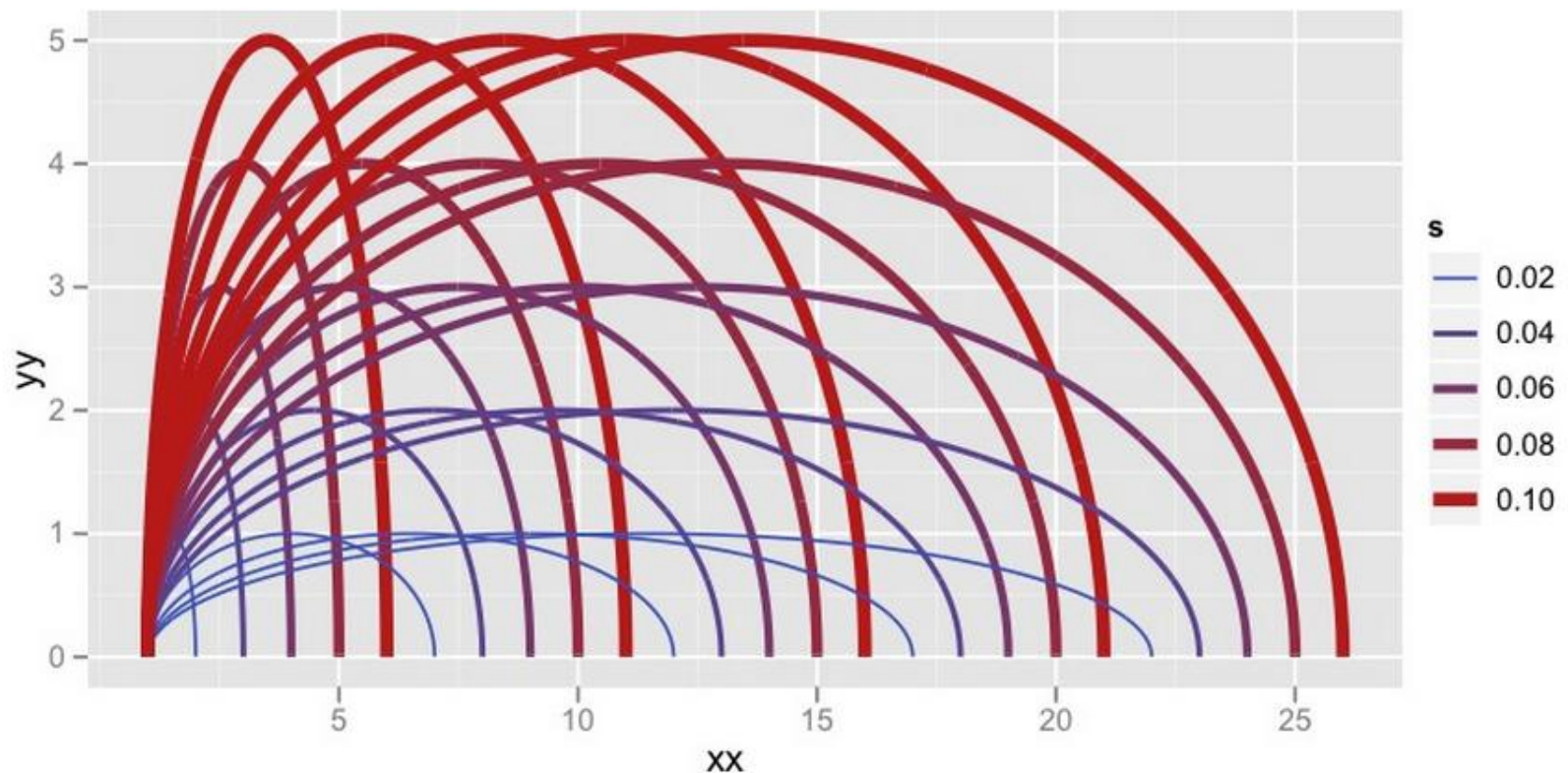
Development:



Notes:

white lines?, color should be in black and white, line width

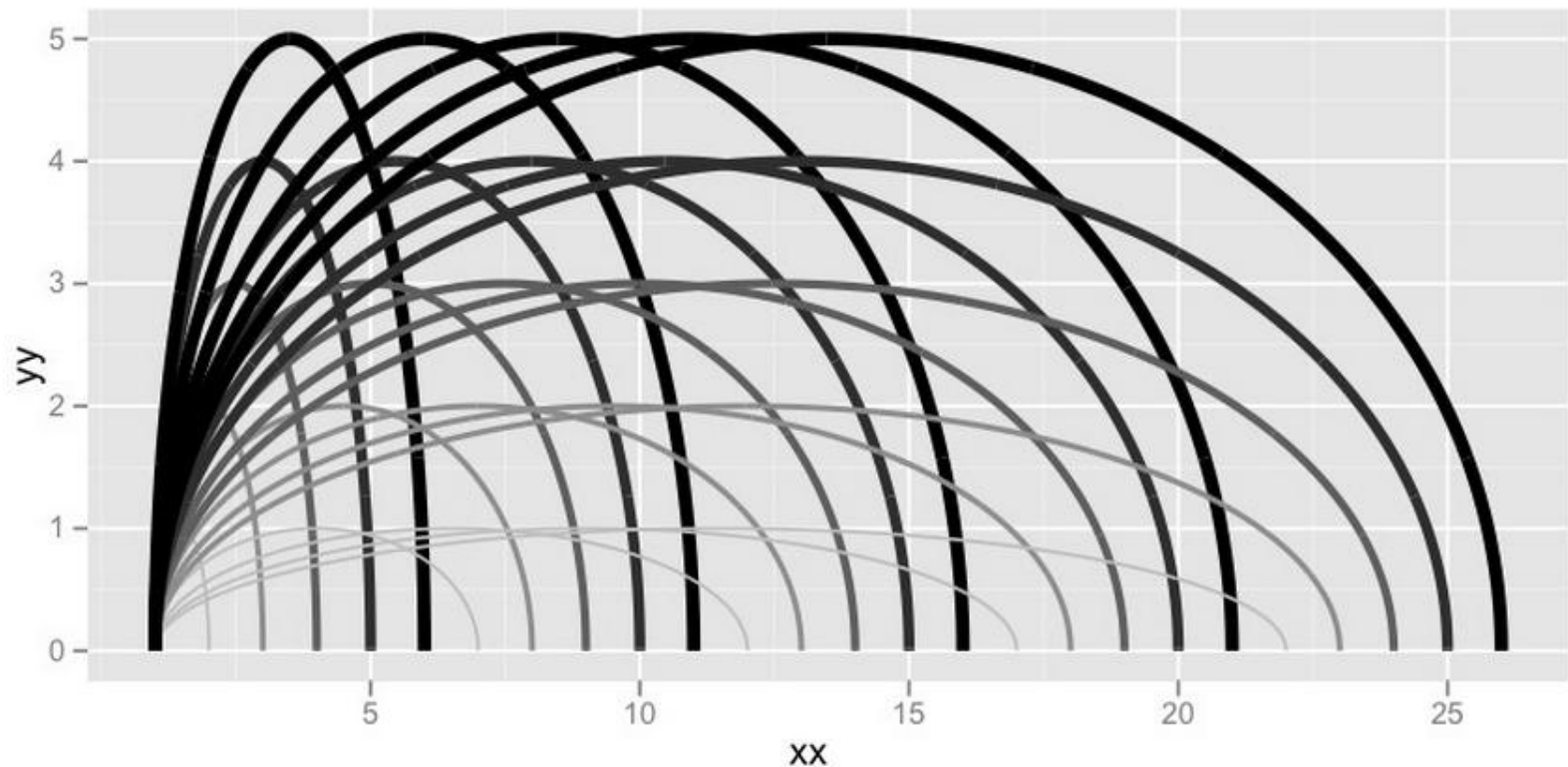
Development:



Notes:

white lines, color should be in black and white

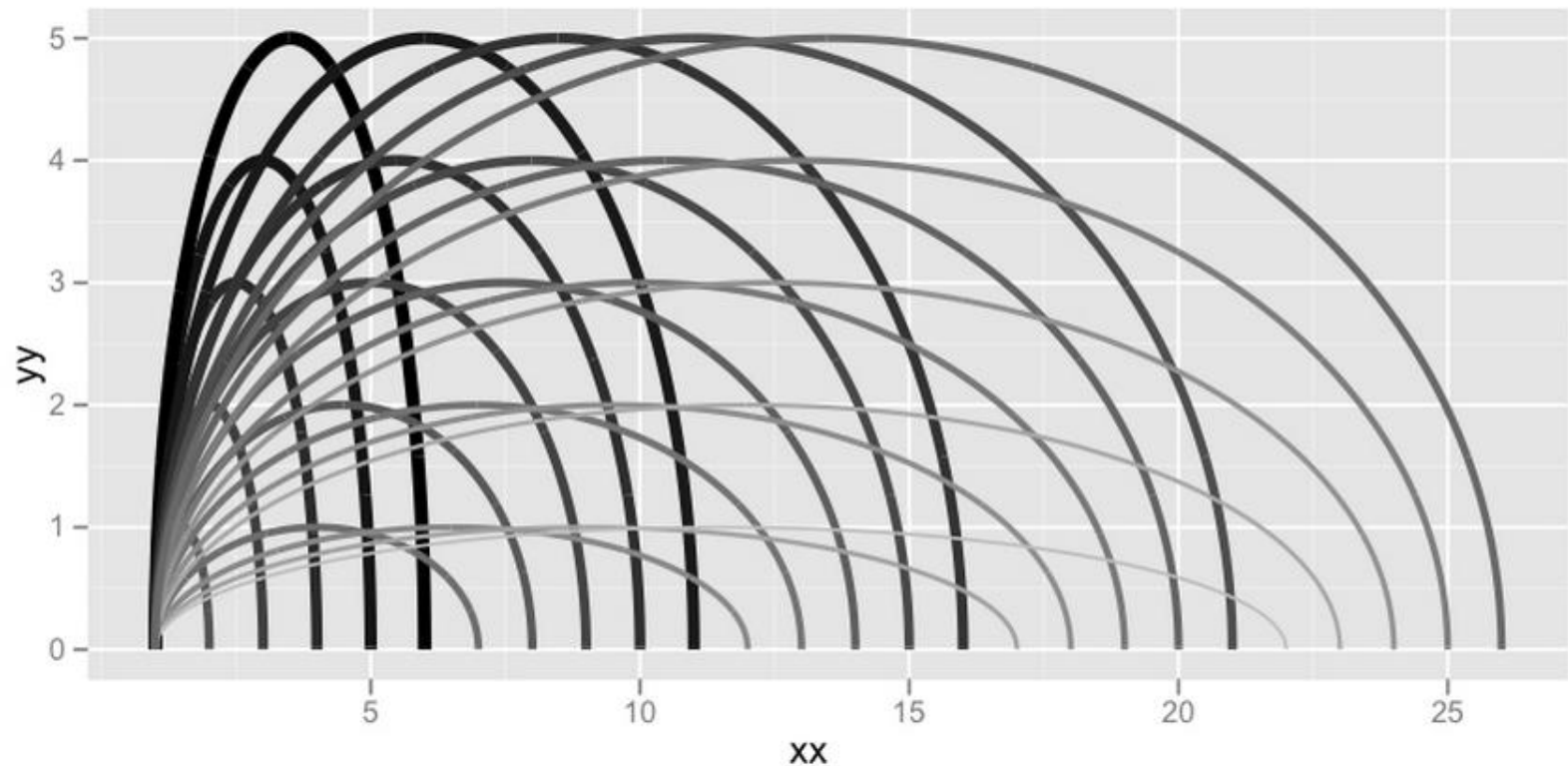
Development:



Notes:

lines changed to black and white but not within
geom_arch function

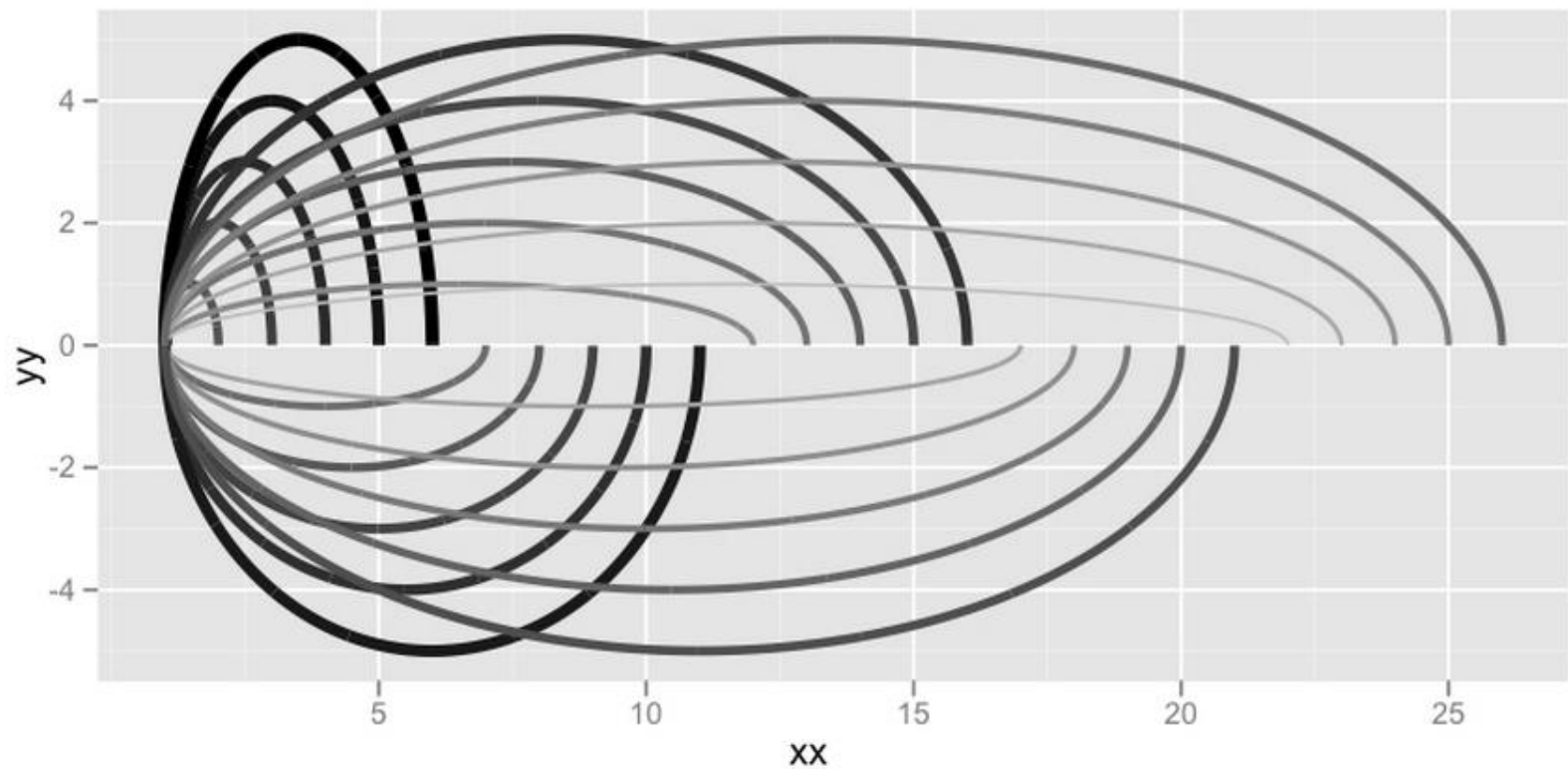
Development:



Notes:

Jump width and height both influence line width and shading

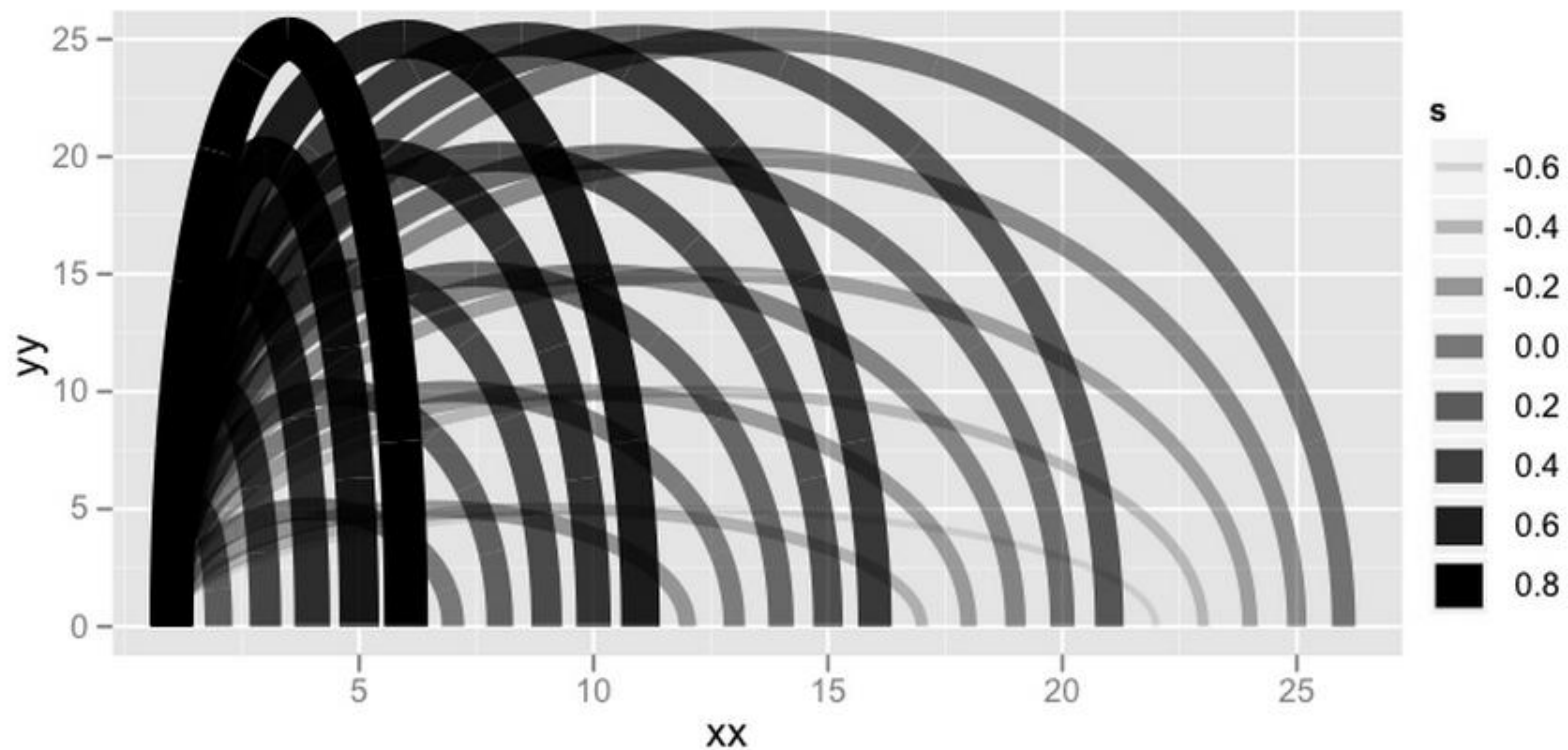
Development:



Notes:

Can display multiple groups of junction reads by flipping one set around the x axis.

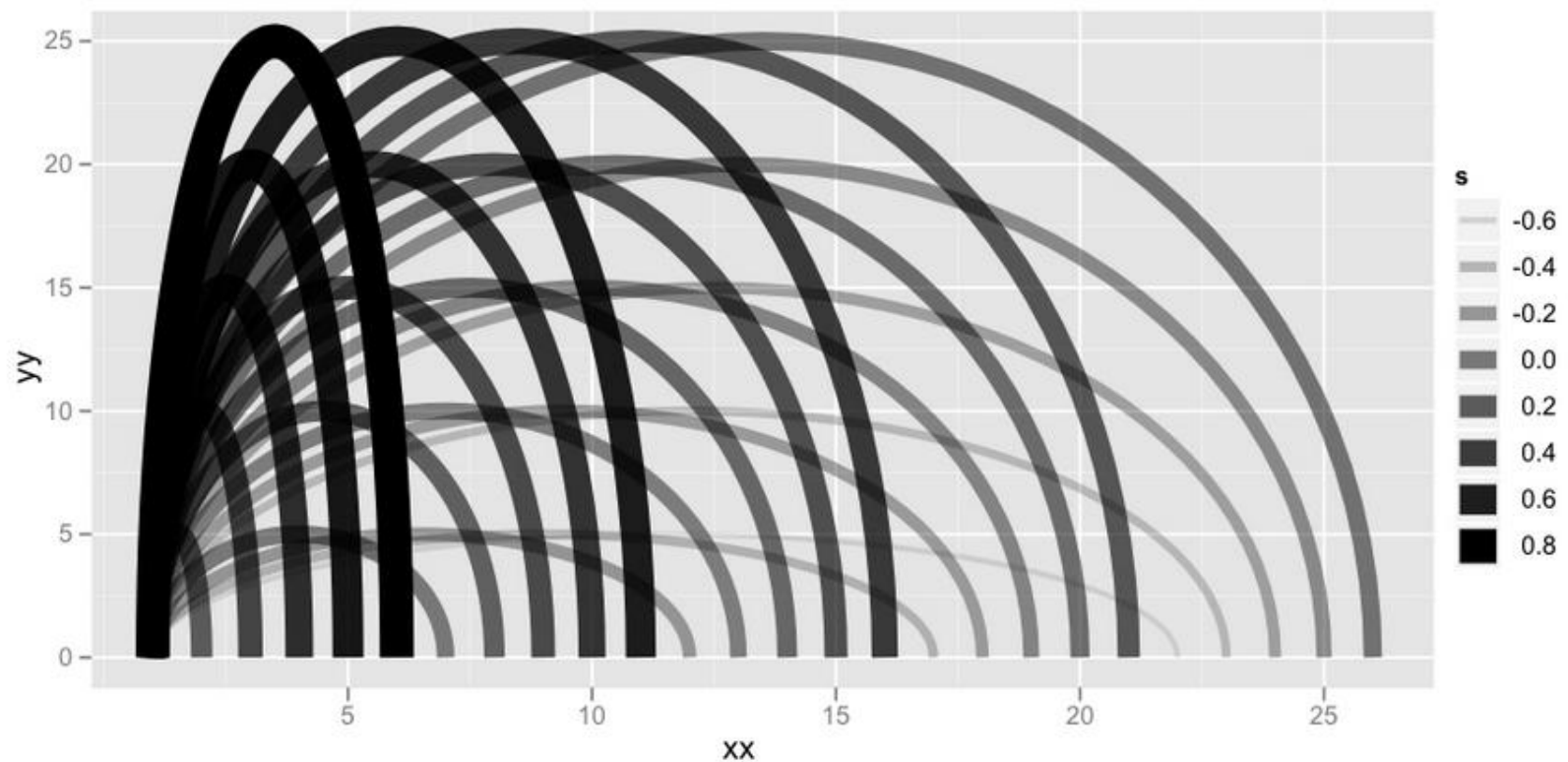
Development:



Notes:

Changed to using alpha instead of shades of grey. Still problems with white lines.

Development:

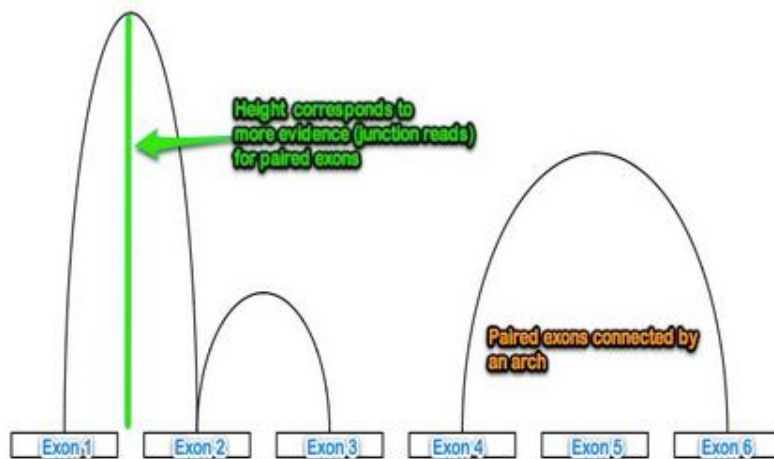


Notes:

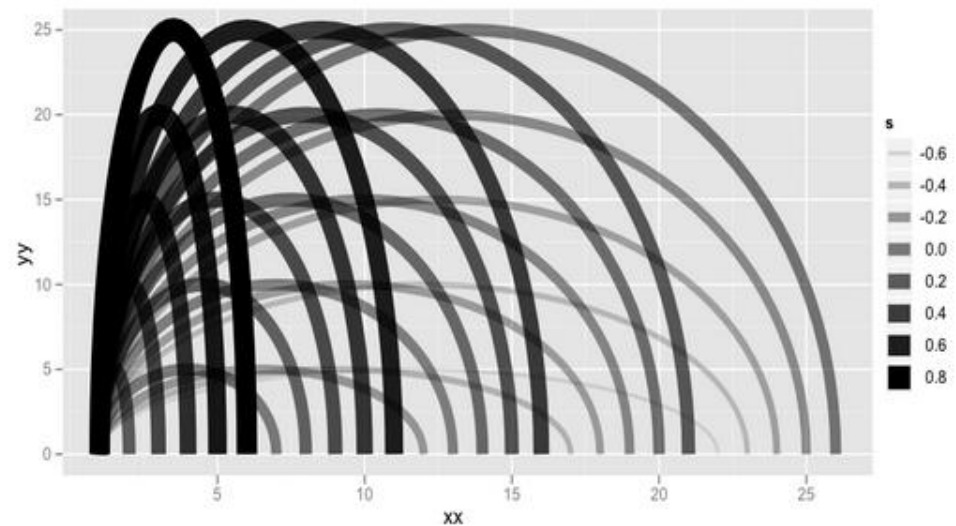
Fixed the white line breaks by plotting less points when creating arch.

Conclusions:

Concept



Completion (so far)



Command:

```
geom_arch(data = , aes(x = , xend = , y = , height = ,))
```

WordPressSite: <http://jcrotone.wordpress.com/>

Current State & Future Plans

bioconductor/ggbio
Tengfei Yin

[Home](#)[Install](#)[Help](#)

[Home](#) » [Bioconductor 2.11](#) » [Software Packages](#) » ggbio

ggbio

Static visualization for genomic data.

Bioconductor version: Release (2.11)

The ggbio package extends and specializes the grammar of graphics for biological data. The graphics are designed to answer common scientific questions, in particular those often asked of high throughput genomics data. All core Bioconductor data structures are supported, where appropriate. The package supports detailed views of particular genomic regions, as well as genome-wide overviews. Supported overviews include ideograms and grand linear views. High-level plots include sequence fragment length, edge-linked interval to data view, mismatch pileup, and several splicing summaries.

Author: Tengfei Yin, Dianne Cook, Michael Lawrence

Maintainer: Tengfei Yin <yintengfei at gmail.com>

To install this package, start R and enter:

```
source("http://bioconductor.org/biocLite.R")
biocLite("ggbio")
```

To cite this package in a publication, start R and enter:

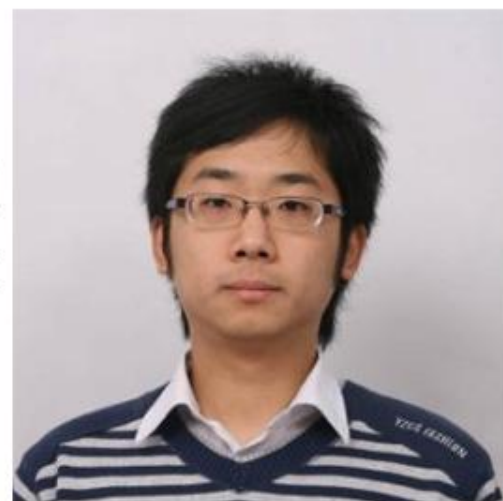
```
citation("ggbio")
```


Acknowledgements: (Thank You!)



Dr Di Cook
Department of Statistics

Tengfei Yin
PhD Student
Biomathematics & BCB



Michael Lawrence
(Not pictured) at Genentech

Major Adviser: Dr Hui-Hsien Chou
Complex Computational Laboratory