

Data Visualisation in Data Science
Instructions for retake project

Prof Jan Aerts, Jelmer Bot

I. General idea

For the integrated project, we ask you to develop a sveltekit application that includes different visuals. More specifically, you will create a hive plot (see the introductory slides, and <http://www.hiveplot.com/>) as well as the overview of the genes on the E.coli genome.

IMPORTANT

This is an **individual** project.

I.I. The data

The E.coli genome consists of a single circular chromosome. Each gene has a start and stop position on that chromosome, as well as a "strand". This strand indicates in which direction the gene points: clockwise or counter-clockwise.

I.I.I. Gene information

For all genes in the E.coli genome, you will get information like the following:

Sample document for gene

```
{
  "gene_id": "ECK120001066",
  "gene_name": "nhaR",
  "gene_blattnr_number": "b0020",
  "gene_start": 18715.0,
  "gene_stop": 19620.0,
  "gene_strand": "forward",
  "gene_product": "DNA-binding transcriptional activator NhaR",
  "gene_evidence": null,
  "gene_pmids": "1316901,1630901,2429258,3413113,8168494,",
  "gene_confidence": null,
  "gene_all_blattnr": "b0020",
  "gene_other_ids": null,
  "ngn": "nhar",
  "axes": "manager"
}
```

Some important elements that you'll need in the assignment:

- **gene_id**: you'll need this to link to an external website
- **gene_start**: you'll need this to draw the gene in the "gene overview" visual
- **gene_strand**: you'll need this to draw the gene in the "gene overview" visual
- **ngn**: stands for normalised gene name, and is the unique gene name you can use across the different views
- **axes**: This value can be empty (if not involved in any interaction), or one of the terms mentioned in the picture below ("Hiveplot example by Martin Krzywinski")
 - **regulator**: gene that regulate other genes, but are themselves not regulated
 - **workhorse**: gene that is regulated by other genes, but itself does not regulate others
 - **manager**: gene that does both

The URL for this dataset is <https://vda-lab.github.io/assets/genes.json>

1.1.2. Interactions

For all gene-gene interactions, you will get information like the following:

Sample document for interaction

```
{
  "from_id": "ECK120011186",
  "from_name": "Fis",
  "from_gene_name": "fis",
  "to_id": "ECK125286587",
  "to_name": "CRP-Sxy",
  "function": "-",
  "confidence": "Strong",
  "from_ngn": "fis",
  "to_ngn": "crpsxy"
}
```

NOTE

Not all genes in the `genes.json` file are involved in interactions!

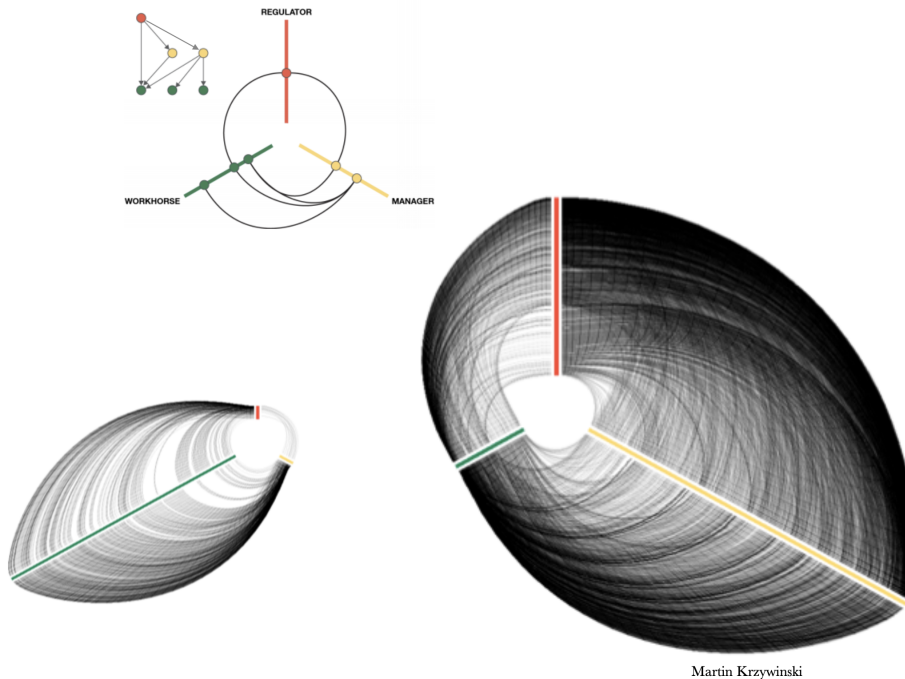


Figure 1. Hiveplot example by Martin Krzywinski

The URL for this dataset is <https://vda-lab.github.io/assets/interactions.json>

NOTE

You can choose to load the data directly from these websites, or download them within your project. The choice is yours.

2. Setting up

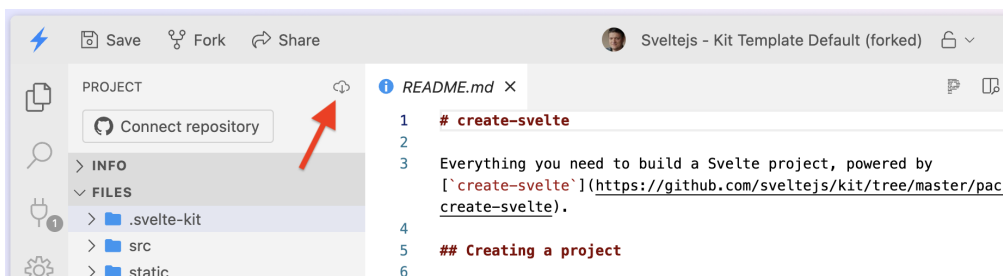
You can either work on your own machine, or use stackblitz. In either case, what you will submit to us will be a zip-file of your code base (see below).

When developing on your own machine, see the sveltekit website (kit.svelte.dev) for instructions on how to set things up. At the time of writing this document, this means running `npm create svelte@latest my-app`, but please check yourself.

When using stackblitz, we will *not* provide you with a template. At the time of writing this document, you can create a new sveltekit application by going to <http://node.new/sveltekit>. This will create a demo application instead of an empty shell, so you will have to remove some files. These are (among others):

- all files as well as the `about` directory underneath the `routes` folder
- the `lib/images` folder

To download your stackblitz codebase as a zip-file, use the "Download Project" button as shown in the figure below:



NOTE

We cannot guarantee that these instructions are still correct when you start your project. So please set things up as soon as possible.

IMPORTANT

You are responsible for your code. Take precautions so that you do not accidentally loose/delete it.

3. Storyboard

Below we describe the specific visuals and interactions that we request from you. In the "Evaluation" section, we will describe very specifically what components you will be graded on.

This image gives an overview of the different pages in the application and how they are linked to each other:

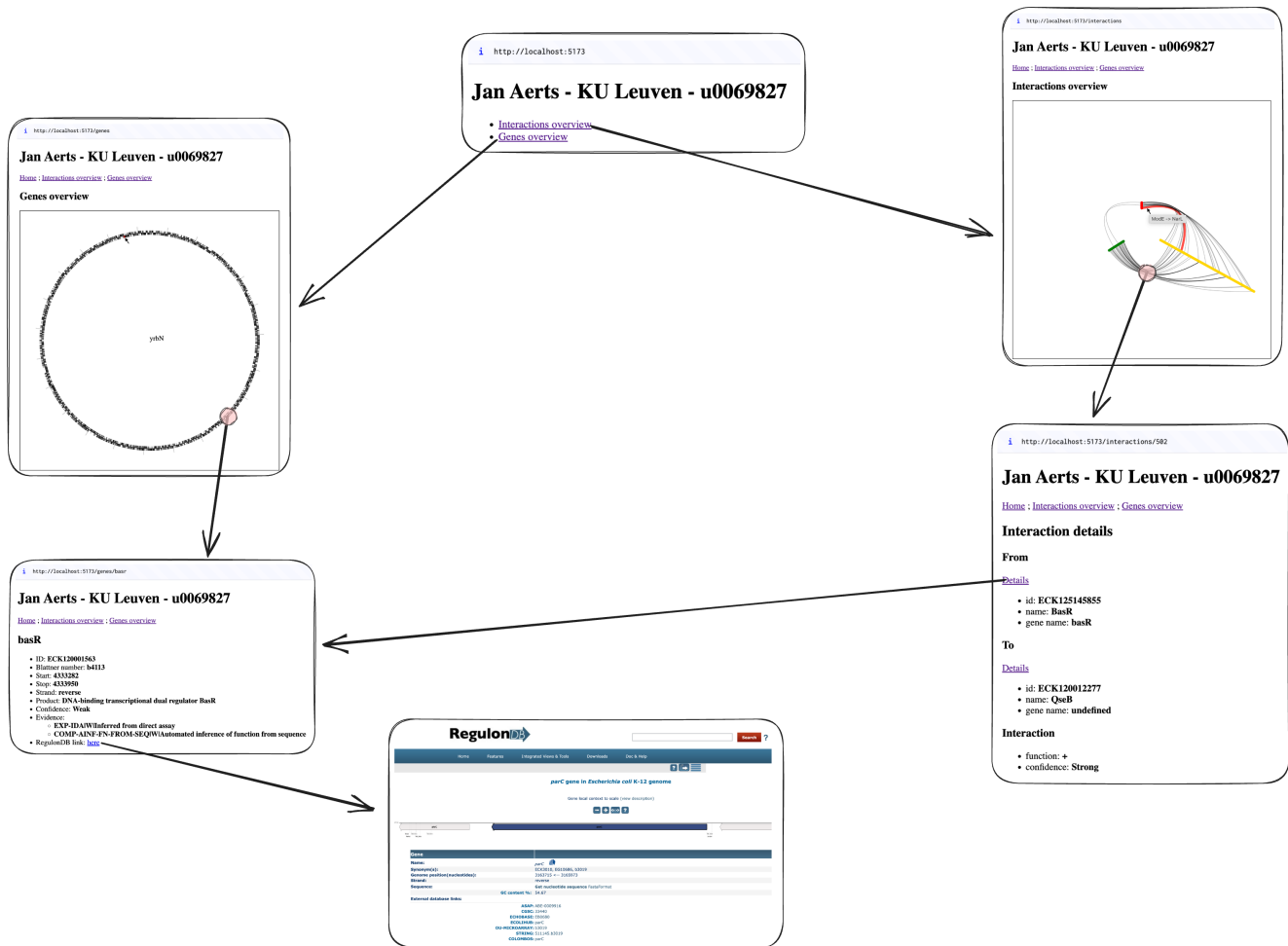


Figure 2. General overview of the application

3.1. Overview page: general links

As shown in the figure, you should create a page showing:

- Your name, university and student number
- Links to the genes and interactions overview pages

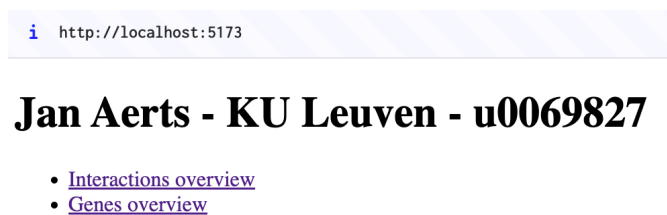


Figure 3. Main page

3.2. Overview page: genes

The genes overview page should contain your name, university and student number, as well as a visual representing the chromosome. Below are the specifications for this visual:

Specifications for visual encoding:

- The SVG has a size of 600x600 pixels.
- The circle has a radius of 250px.
- The start of each gene is indicated as a small line on that circle.
- For genes on the **forward** strand, this should be on the *outside* of the circle; genes on the **reverse** strand should be on the *inside* of the circle.
- Genes should have an opacity of 50%
- Lighter, longer lines indicate ticks along the chromosome. It is not necessary to add the actual value to those tick marks. Ticks should be added for every 200,000 positions.

When hovering over a gene, the following should happen:

- That gene line gets a higher opacity.
- It is red instead of black;
- The name of the gene is shown in the middle of the circle.

When clicking on a gene, the following should happen:

- The gene details page is opened.

 <http://localhost:5173/genes>

Jan Aerts - KU Leuven - u0069827

[Home](#) ; [Interactions overview](#) ; [Genes overview](#)

Genes overview

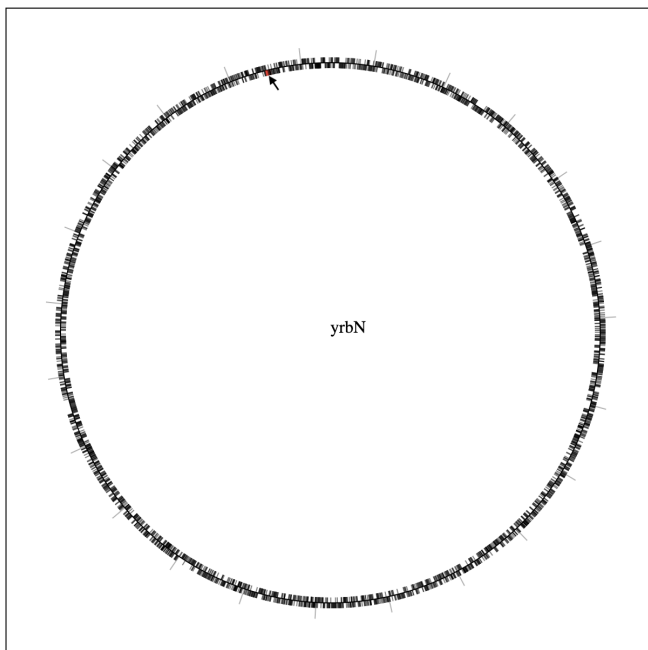


Figure 4. Gene overview with interaction

3.3. Details page: gene

The gene details page should contain your name, university and student number, as well as a details for that gene. Below are the specifications:

- The name of the gene (**basR** in the screenshot below) should be a header of level 2.
- There should be an unordered list with the following information:
 - ID
 - Blattner number
 - Start
 - Stop
 - Strand
 - Product
 - Confidence
 - Evidence: is itself an unordered list
 - RegulonDB link: should link to the following URL:
<http://regulondb.ccg.unam.mx/gene?organism=ECK12&format=jsp&type=gene&term=xxxxxx> where **xxxxxx** corresponds to the ID of this gene.

 <http://localhost:5173/genes/basR>

Jan Aerts - KU Leuven - u0069827

[Home](#) ; [Interactions overview](#) ; [Genes overview](#)

basR

- ID: **ECK120001563**
- Blattner number: **b4113**
- Start: **4333282**
- Stop: **4333950**
- Strand: **reverse**
- Product: **DNA-binding transcriptional dual regulator BasR**
- Confidence: **Weak**
- Evidence:
 - **EXP-IDA|W**Inferred from direct assay
 - **COMP-AINF-FN-FROM-SEQ|W**Automated inference of function from sequence
- RegulonDB link: [here](#)

Figure 5. Gene details view

For your information, the RegulonDB website will show something like the page below:



Jan Aerts - KU Leuven - u0069827

[Home](#) ; [Interactions overview](#) ; [Genes overview](#)

Interactions overview

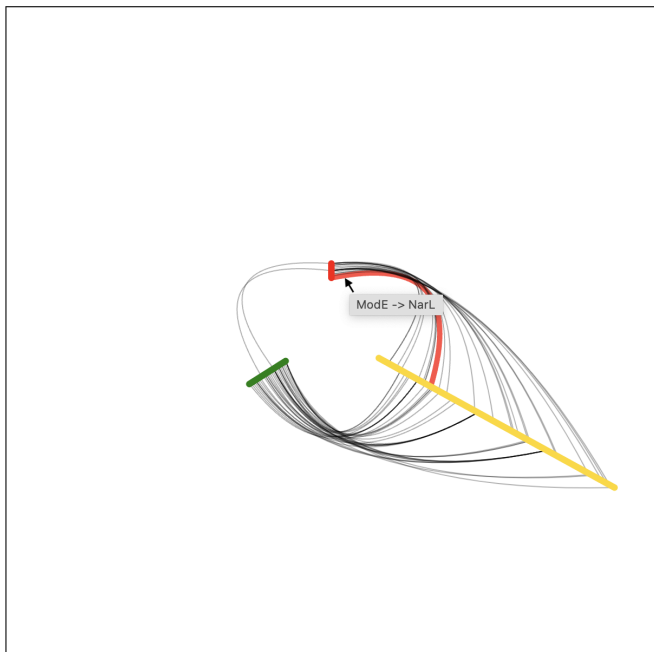


Figure 7. Interactions overview

In a simpler version ("simplification 1"), you can use straight lines instead of Bezier curves for the links. In that case your visual will look like this:

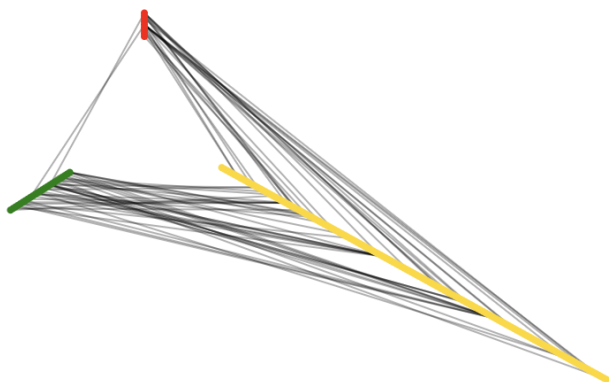


Figure 8. Straight lines instead of Bezier curves

Further simplifying ("simplification 2"), you can decide to not rotate the axes:

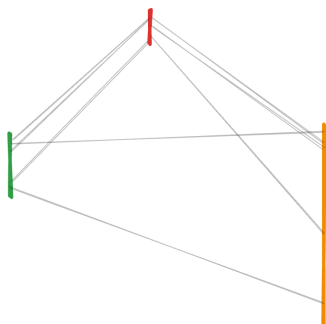


Figure 9. Hive plot without rotated axes

3.5. Details page: interaction

The interaction details page should contain your name, university and student number, as well as a details for that interaction.

For both the `from` and `to` gene, display the `id`, `name` and `gene name`. In addition, list the `function` and `confidence` for the interaction itself.

For both the `from` and `to` gene, also add a `Details` link that points to the gene detail page.



Figure 10. Interaction details

4. Evaluation

You will receive points for the different aspects of this project:

- Each page:
 - has your name, university, student number: 0.25
 - has a link to the main page, the gene overview, and the interactions overview: 0.25
- In the overview page:
 - working link to the gene overview page: 0.5
 - working link to the interactions overview page: 0.5
- In the genes overview page:
 - genes present on the circle: 1.5
 - genes on forward/reverse strand on outside/inside of circle: 1
 - ticks added: 1
 - gene line turns red and bold on hover: 0.5
 - gene name is shown in middle of circle on hover: 1
 - gene details page is opened when clicking on a gene: 1
- In the gene details page:
 - information as specified above is present on page: 0.5
 - RegulonDB link works: 1
- In the interactions overview page
 - genes are organised in lines according to their role (`regulator,manager,workhorse`): 1.5
 - axes are rotated (instead of "simplification_2"): 1.5
 - **only if not simplification_2**: lines are Bezier curves (instead of "simplification_1"): 1
 - lines turn red and bold on hover: 0.5
 - interaction details page is opened when clicking on an interaction: 1
- In the interaction details page:
 - information as specified above is present on the page: 0.5
 - working links to gene detail pages: 1

This amounts to 16/20. The other 4/20 will be based on the designs you submitted. In case no designs were submitted, your maximum score will therefore be 16/20.

4.1. What to hand in?

An assignment will be created on Toledo/Blackboard where you will submit 2 files:

- a **zip-file** containing the code in your repository.
 - Please remove the contents of the `node_modules` directory before zipping the repository.
 - We should be able to unzip the file, run `npm install` (which installs all dependencies), and `npm run dev` to get your interface.
- a short **video** (no sound; in `.mov` or `.mp4` format) of max 30 seconds and max 15MB showing the following (in this order):

1. on the overview page:
 - a. You click on the "gene overview" link
2. on the gene overview page:
 - a. You hover over different genes to show that (a) that gene's name is shown in the center, and (b) that gene's line becomes red and more thick.
 - b. You click on a gene, which should bring you to:
3. on the gene details page:
 - a. You click on the "interactions overview" link
4. on the interactions overview page:
 - a. You hover over several lines to show that that line becomes red and more thick.
 - b. You click on one of those lines, which should bring you to:
5. on the interactions details page:
 - a. You click on one of the details links which should bring you to a RegulonDB webpage for that gene.

A video showing what we expect is available on Toledo/Blackboard.

5. Interaction with the teaching team

The teaching team is of course still available for questions. These should however be focussing on clarification of the assignment.

For technical questions, please check the teaching material at <https://vda-lab.gitlab.io/datavis-technologies/>, the exercises, and google; do not ask the teaching team. There is **one** exception: data loading, because that is a prerequisite for all the aspects of the evaluation. If you need us to help you load the data itself, we will do so but there will be 1 point subtracted from your score.