

# Safety Assessment on Roslagsbanan Rail System, 2021-2022

This manuscript ([permalink](#)) was automatically generated from [uiceds/cee-492-term-project-fall-2022-time@05606a3](#) on October 30, 2022.

## Authors

---

- **Cecilia Karina Volpe Baridon**  
· [cvolpebaridon](#)
- **Elie Roudiere**  
· [elieroudiere](#)  
Railway Group, KTH Royal Institute of Technology - Stockholm, Sweden
- **Berkhan Usta**  
· [berkanusta](#)

## Dataset Description

The Roslagsbanan is a railway system operating in Stockholm. As part of an improvement in the transportation system, Stockholm Municipality invested in a new fleet of trains switching from X10p trains to the new X15p. In order to analyze the safety of the new trains, brake tests were conducted. During these tests, the Latronix Track Measurement (LTM) system was mounted on the train. The LTM uses laser beams, and an onboard computer to perform measurements over a railroad track while the train drives on the track at normal speed.

The data that will be used in this study was collected using the LTM system affixed to a Roslagsbanan vehicle. While the train was running over the sections, the level of each rail, alignment, curvature, cant, gauge, and twist were measured every 256mm. The data was compiled in csv files and was collected between October 2021 and May 2022 with an approximate interval of one month. In other words, a given section of track typically has around six measurement files. The total amount of data is 60 CSV files of 30,000 to 80,000 rows each, for a total of 1.5GB of data.

*These data and measurements are property of Latronix AB of Sweden, and are used with their permission for research and educational purposes only.*

In the CSV files the structure of columns is as follows:

- **Marker and Marker\_Offset:** The kilometre post of the data point, where Marker is in km and Marker\_offset is in m. In other words, if Marker=20 and Marker\_offset=250.5, the kilometre post is 20,250.5
- **Coordinate\_northing, Coordinate\_easting, Latitude, Longitude:** As their name suggests, they are the coordinates of the data point in northings and eastings, as well as latitude and longitude.
- **Speed:** records the speed in km/h at which the measurement was taken. This is especially relevant as different operating speeds give different tolerances in terms of the measured properties.

- Level\_left and Level\_right D0, D1, D2, D3: The deviation of the longitudinal level (height) of the left and right rail in mm, compared to its reference position. D0, D1, D2, D3 correspond to different measurement wavelengths of the same data.
- Alignment\_left and Alignment\_right D0, D1, D2, D3: The deviation in longitudinal alignment ("straightness") of the left and right rail in mm, compared to its reference position. D0, D1, D2, D3 correspond to different measurement wavelengths of the same data.
- Track\_gauge: The deviation in the spacing between the rails in mm, from a reference of 891mm.
- Curvature: The curvature of the track at that location, in  $m^{-1}$ .
- Cant: The cant (inclination) of the track at the location, in mm from the horizontal.
- Twist\_3m, Twist\_6m: The rate of change in cant over a 3m or 6m distance.
- Level\_versine\_left\_5m, Level\_versine\_right\_5m: Derived from Level\_left and Level\_right, this level deviation in mm corresponds to longitudinal versine measurement of the top of rail over a 5m distance. This historical metric is more commonly used in railway engineering and is thus provided.
- Alignment\_versine\_left\_10m, Alignment\_versine\_right\_10m: Similarly to the versine level, this measure of alignment in mm uses versines over a 10m distance.

Finally, Track\_name, Track\_class, Station\_flag and Link\_name, Pos\_unfiltered Pos\_filtered, NTP\_sync, Cant\_D1, Std\_level, Std\_alignment, Std\_cant, Quality\_1 are mainly for internal use and not part of the analysis.

## Project Proposal

The main goal of this project is to identify patterns in the data, especially the track degradation of certain sections, and thus to analyze the change in railway safety during this time. The first approach will be to define degradation trends and build a model to process the data that finds them automatically. Then, the study will attempt to find external factors which caused the observed trends. The analysis will include graphics of the parameters over time as well as comparing them over the sections to provide a better understanding of the data visually.

This research will help assess system performance, identify possible maintenance schedules, and provide useful data for further research in this area. The study also falls within the scope of developing predictive maintenance in civil engineering and railways. Creating and improving analysis and predictive maintenance models such as this one can contribute to increase safety of rail networks and a higher quality of service for passengers. They will benefit from fewer unplanned service interruptions due to track maintenance and even fewer unsafe tracks causing line closures.

## Exploratory Analysis

### General Dataset Description:

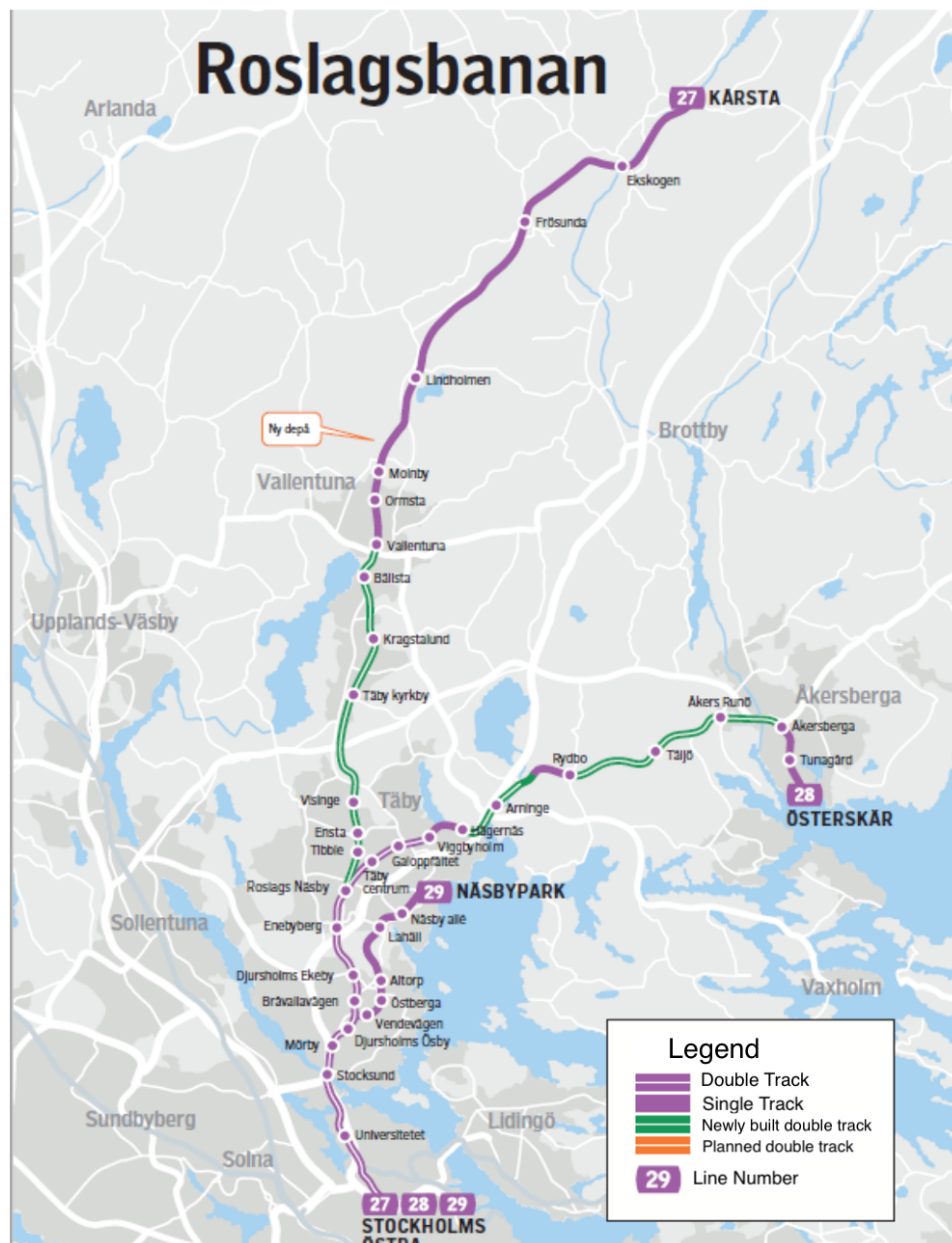
---

The main dataset is composed of 60 .csv files of measurements taken on part of the railway tracks of the Roslagsbanan network in Stockholm. They are dated from October 2021 to May 2022 and the

measurements were made using track measuring equipment installed on the trains and working during the trial run.

Each file corresponds to a specific measurement run, on a given part of the track. Additionally, each section of the track has a different number of measurement runs, depending on the importance of the sections. For example, the segment between the 27 and 38 km points comprises more than 20 different runs (usually traveling north and then south on the same day) while other sections do not have such a number of measurements. Figure 1 shows a map with the lines that are taken into account in this document. (we should cite all the Figures in the text, I don't know how to do that automatic)

`. It looks interesting to analyse`



**Figure 1:** Line map overview of the Roslagsbanan. This study will focus on line 27 to Kårsta.

It should be noted that in some cases the data needs to be cleaned. For example, it contains many N/A cells, as well as some questionable data points, especially when the train was coming to a stop, which produced some outliers. [plot showing what happens with 0 speed?]. I like this idea too

The date and time of the measurements are not hardcoded in the .csv file but only mentioned in the file name, as such the code for analyzing will have to take into account the name of the file when making comparisons. As shown in Figure 2 , the measurement runs are not completely synchronized, they must match or correlate in some way to clearly interpret the results. (WHAT IS X AND Y IN THE FIGURE?). I think that the letter is too small, maybe we could make it bigger

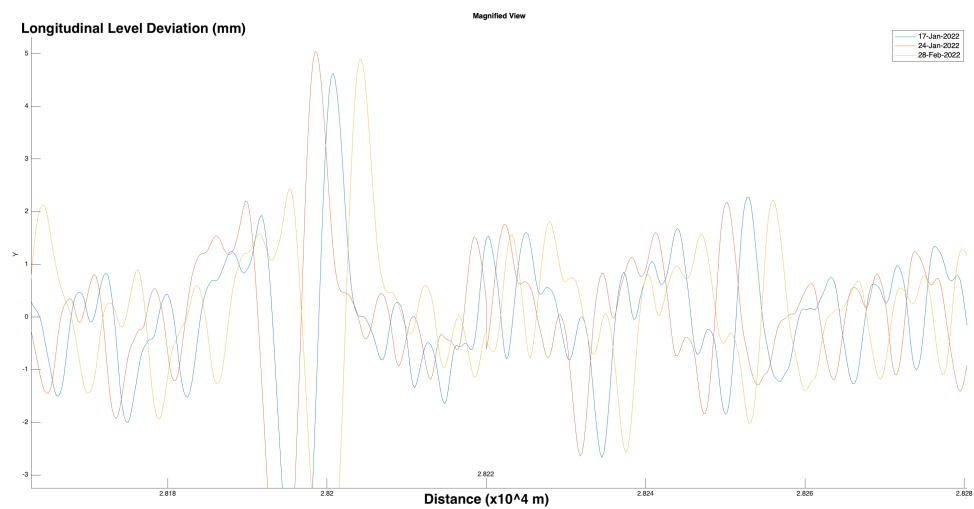


Figure 2: Plot showing the offset between measurement runs on three dates.

## Statistics Analysis:

[Stats summary about size, distance covered, frequency of measurements, others?]

Statistic	Value
Earliest measurement	21st October 2021
Latest measurement	10th May 2022
Line Length	41.700 km (25.91 miles)
Data points per file	20,000 to over 100,000
File size	

Table 1: Summary Table. {#tbl:sum-table-1}

Something is weird with the title of this table, it appears something else in the pdf version We could include in this section the mean and standard deviation of the columns, for instance, gauge, alignment, height (that wee know they should be the same in the entire longitude)

## Data Correlation:

An important fact of the parameters considered is that some of them are directly related in a mathematical sense. For example, the parameter cant is proportional to the curvature, which means that when the track is being built, the superelevation is calculated based on the curvature at a given location, as follows:

()

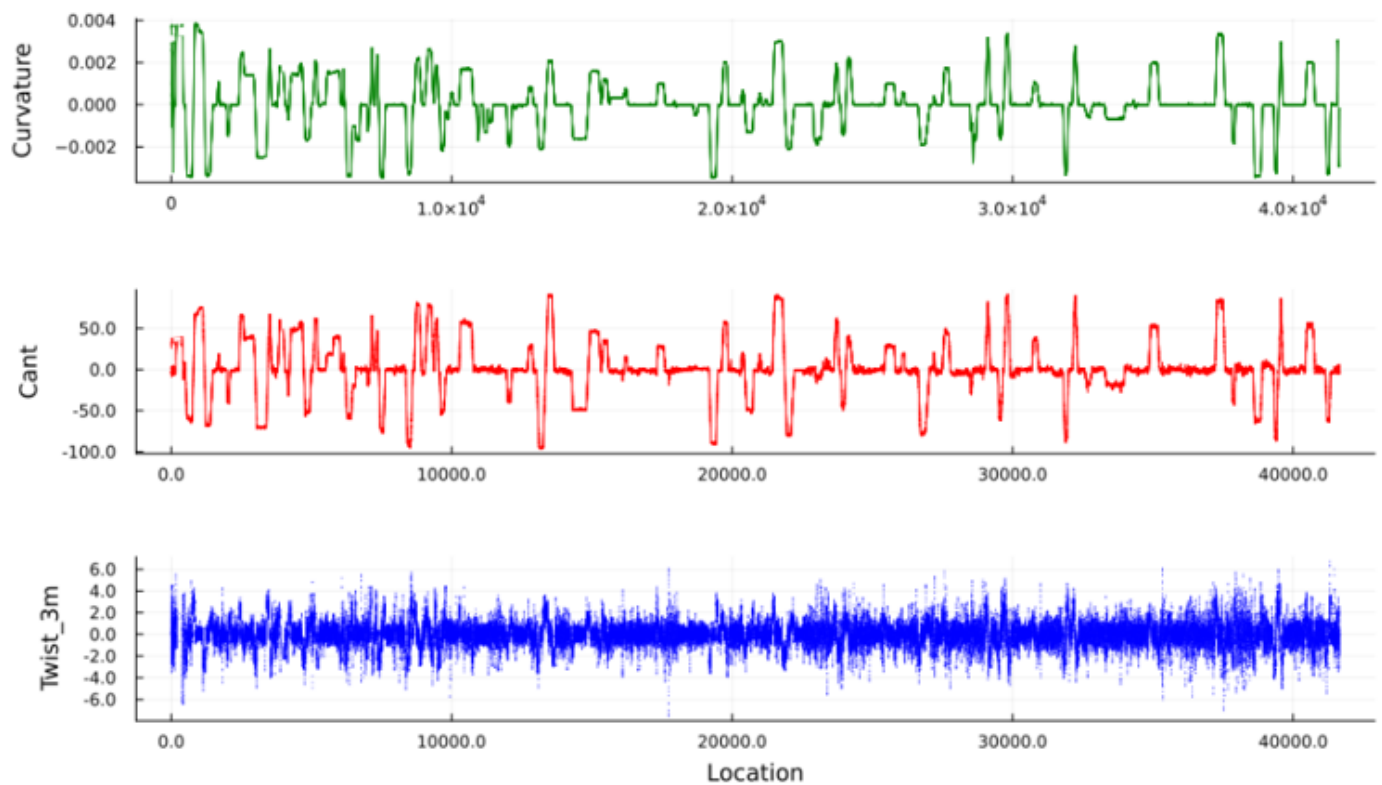
Where  $E_e$  is balancing superelevation or cant,  $G$  (track gauge),  $g$  (gravity), and  $V$  (line speed) are constants and  $R$  is curve radius, the inverse of curvature.[1]

Similarly, twist describes the rate of change of cant over a certain distance, and thus a close correlation between them is expected. These parameters are summarised in Figure 3 below, courtesy of Kawasaki Track Technology:



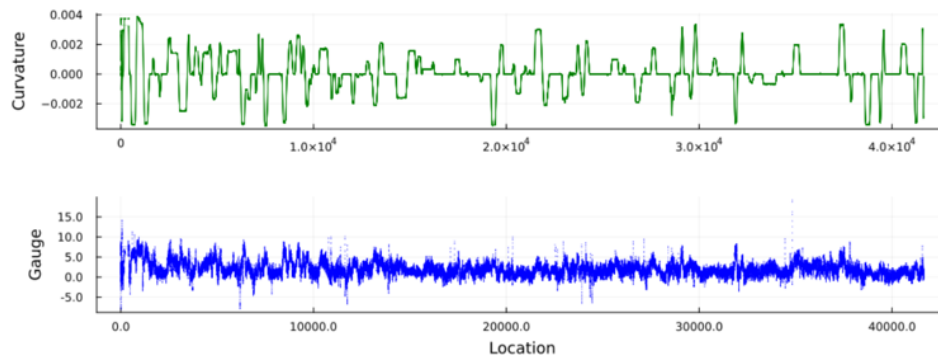
**Figure 3:** Summary diagram of the parameters studied in this report. Source: Kawasaki Track Technology

Figure 4 shows these relations among the parameters for the Karsta line. From the curvature and cant plots, it can be observed that there is a strong relationship between them, as both exhibit similar patterns. However, in the case of twist correlation with the other variables, similar patterns may not be very clear. More analysis might be done on this line.

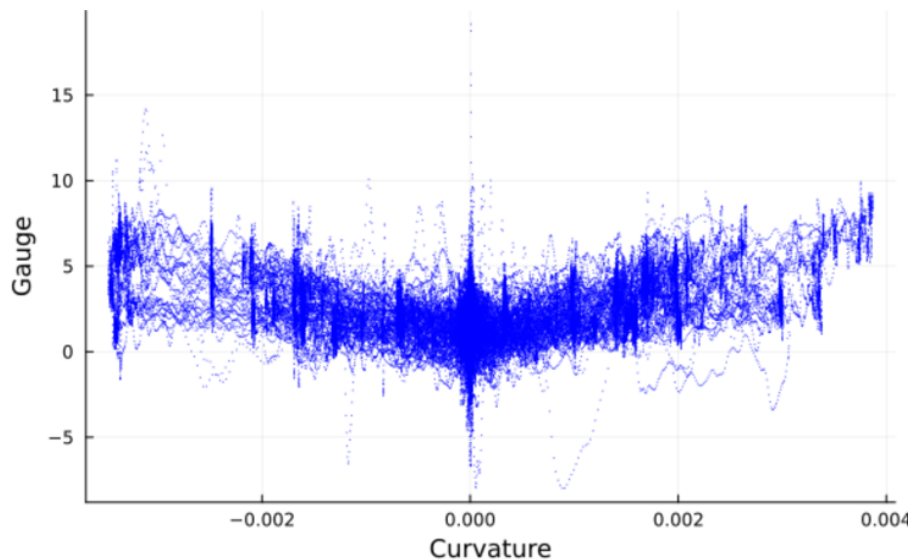


**Figure 4: Ostra-Karsta line - Curvature, Cant and Twist vs Location.**

As for Gauge and Curvature, while there is no strict relationship between them when plotting the two parameters, it can be observed that at higher curvatures, the distance between tracks also increases. Figure 5 and Figure 6 show this fact.



**Figure 5: Ostra-Karsta line - Curvature and Gauge vs Location.**



**Figure 6: Ostra-Karsta line - Curvature vs Gauge.**

## Predictive modelling

Due to the somewhat limited amount of independent variables already present inside the measurement set, there may be challenges. Correlation plots can help us determine links between variables and what model to create.

The first part of the predictive model will try and find locations where the track degradation seems to form a clear trend (typically a linear increase in a given direction) over time. Here, time is the independent variable. From an initial analysis of the dataset, it is likely this predictive model will yield some results (picture of a steady trend?). The second part of the predictive model we would like to create would take into account external factors, in particular temperature/weather changes and soil properties if obtainable. This could give us locations prone to developing trends in the short to medium term, which could be investigated by a maintenance team. However we have envisioned some possible issues, namely that obtaining and translating soil data into something usable by the model may be a challenge, and no link is guaranteed to appear (what if all soils are similar?). A possible approach could be to grade soils based on properties, for example a very stable soil could get a score of 1 and a worse soil from the point of view of building a railway line could get a score of 4 (or higher/lower), with appropriate intermediate values. We could look into D30/D10/D60 values.

This manuscript is a template (aka “rootstock”) for [Manubot](#), a tool for writing scholarly manuscripts. Use this template as a starting point for your manuscript.

The rest of this document is a full list of formatting elements/features supported by Manubot. Compare the input ( `.md` files in the `/content` directory) to the output you see below.

# Basic formatting

---

## Bold text

## Semi-bold text

Centered text

Right-aligned text

*Italic text*

Combined *italics and* **bold**

~~Strikethrough~~

1. Ordered list item
2. Ordered list item
  - a. Sub-item
  - b. Sub-item
    - i. Sub-sub-item
3. Ordered list item
  - a. Sub-item

- List item
- List item
- List item

subscript: H<sub>2</sub>O is a liquid

superscript: 2<sup>10</sup> is 1024.

[unicode superscripts](#)<sup>0123456789</sup>

[unicode subscripts](#)<sub>0123456789</sub>

A long paragraph of text. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Putting each sentence on its own line has numerous benefits with regard to [editing](#) and [version control](#).

Line break without starting a new paragraph by putting two spaces at end of line.

## Document organization

---

Document section headings:

# Heading 1

## Heading 2

---

### Heading 3

#### Heading 4

##### Heading 5

###### Heading 6



**A heading centered on its own printed page**

Horizontal rule:

---

Heading 1's are recommended to be reserved for the title of the manuscript.

Heading 2's are recommended for broad sections such as *Abstract*, *Methods*, *Conclusion*, etc.

Heading 3's and Heading 4's are recommended for sub-sections.

## Links

---

Bare URL link: <https://manubot.org>

[Long link with lots of words and stuff and junk and bleep and blah and stuff and other stuff and more stuff yeah](#)

[Link with text](#)

[Link with hover text](#)

[Link by reference](#)

## Citations

---

Citation by DOI [[1](#)].

Citation by PubMed Central ID [[2](#)].

Citation by PubMed ID [[3](#)].

Citation by Wikidata ID [[4](#)].

Citation by ISBN [[5](#)].

Citation by URL [[6](#)].

Citation by alias [[7](#)].

Multiple citations can be put inside the same set of brackets [[1](#),[5](#),[7](#)]. Manubot plugins provide easier, more convenient visualization of and navigation between citations [[2](#),[3](#),[7](#),[8](#)].

Citation tags (i.e. aliases) can be defined in their own paragraphs using Markdown's reference link syntax:

## Referencing figures, tables, equations

---

Figure [7](#)

Figure [8](#)

Figure [9](#)

Figure [10](#)

Table [1](#)

Equation [1](#)

Equation [2](#)

## Quotes and code

---

Quoted text

Quoted block of text

Two roads diverged in a wood, and I—  
I took the one less traveled by,  
And that has made all the difference.

Code `in the middle` of normal text, aka `inline code`.

Code block with Python syntax highlighting:

```
from manubot.cite.doi import expand_short_doi

def test_expand_short_doi():
    doi = expand_short_doi("10/c3bp")
    # a string too long to fit within page:
    assert doi == "10.25313/2524-2695-2018-3-vliyanie-enhansera-copia-i-
        insulyatora-gypsy-na-sintez-ernk-modifikatsii-hromatina-i-
        svyazyvanie-insulyatornyh-belkov-vtransfetsirovannyh-geneticheskikh-
        konstruktsiyah"
```

Code block with no syntax highlighting:

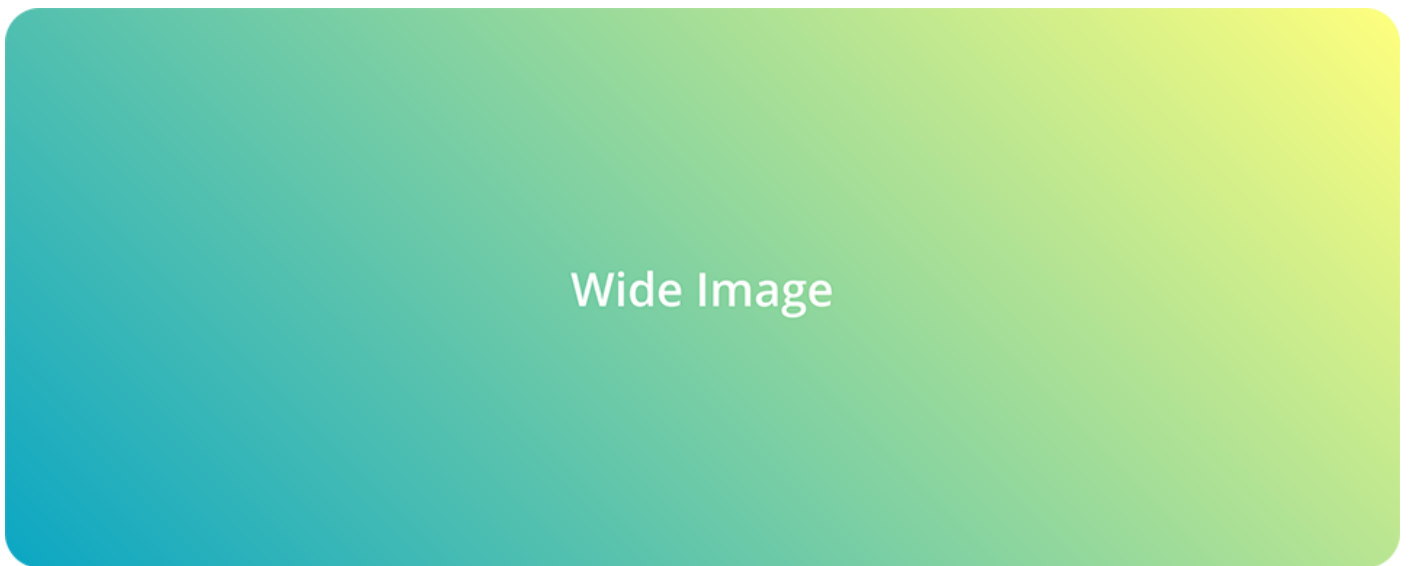
```
Exporting HTML manuscript
Exporting DOCX manuscript
Exporting PDF manuscript
```

## Figures

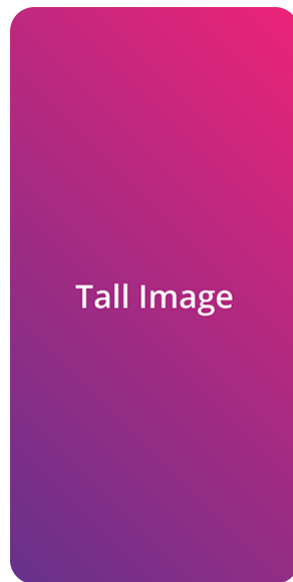
---



**Figure 7: A square image at actual size and with a bottom caption.** Loaded from the latest version of image on GitHub.



**Figure 8: An image too wide to fit within page at full size.** Loaded from a specific (hashed) version of the image on GitHub.



**Figure 9: A tall image with a specified height.** Loaded from a specific (hashed) version of the image on GitHub.



**Figure 10: A vector `.svg` image loaded from GitHub.** The parameter `sanitize=true` is necessary to properly load SVGs hosted via GitHub URLs. White background specified to serve as a backdrop for transparent sections of the image.

## Tables

**Table 1:** A table with a top caption and specified relative column widths.

<i>Bowling Scores</i>	Jane	John	Alice	Bob
Game 1	150	187	210	105
Game 2	98	202	197	102
Game 3	123	180	238	134

**Table 2:** A table too wide to fit within page.

	Digits 1-33	Digits 34-66	Digits 67-99	Ref.
pi	3.14159265358979323846264338327950	288419716939937510582097494459230	781640628620899862803482534211706	<code>piday.org</code>
e	2.71828182845904523536028747135266	249775724709369995957496696762772	407663035354759457138217852516642	<code>nasa.gov</code>

**Table 3:** A table with merged cells using the `attributes` plugin.

	Colors	
Size	Text Color	Background Color
big	blue	orange
small	black	white

## Equations

A LaTeX equation:

$$\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$$

(1)

An equation too long to fit within page:

$$x = a + b + c + d + e + f + g + h + i + j + k + l + m + n + o + p + q + r + s + t + u + v + w + x + y + z + 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9$$

(2)

## Special

**⚠ WARNING** The following features are only supported and intended for `.html` and `.pdf` exports. Journals are not likely to support them, and they may not display correctly when converted to other formats such as `.docx`.

LINK STYLED AS A BUTTON

Adding arbitrary HTML attributes to an element using Pandoc’s attribute syntax:

Manubot Manubot Manubot Manubot Manubot. Manubot Manubot Manubot Manubot. Manubot Manubot Manubot. Manubot Manubot. Manubot.

Adding arbitrary HTML attributes to an element with the Manubot `attributes` plugin (more flexible than Pandoc’s method in terms of which elements you can add attributes to):

Manubot Manubot Manubot Manubot Manubot. Manubot Manubot Manubot Manubot. Manubot Manubot Manubot. Manubot Manubot. Manubot.

Available background colors for text, images, code, banners, etc:

white lightgrey grey darkgrey black lightred lightyellow lightgreen  
lightblue lightpurple red orange yellow green blue purple

Using the [Font Awesome](#) icon set:

✓ ? ★ 🔔 ⛔ …



### Light Grey Banner

useful for *general information* - [manubot.org](https://manubot.org)



### Blue Banner

useful for *important information* - [manubot.org](https://manubot.org)



### Light Red Banner

useful for *warnings* - [manubot.org](https://manubot.org)

- 1. Sci-Hub provides access to nearly all scholarly literature**  
Daniel S Himmelstein, Ariel Rodriguez Romero, Jacob G Levernier, Thomas Anthony Munro, Stephen Reid McLaughlin, Bastian Greshake Tzovaras, Casey S Greene  
*eLife* (2018-03-01) <https://doi.org/ckcj>  
DOI: [10.7554/elife.32822](https://doi.org/10.7554/elife.32822) · PMID: [29424689](https://pubmed.ncbi.nlm.nih.gov/29424689/) · PMCID: [PMC5832410](https://pubmed.ncbi.nlm.nih.gov/PMC5832410/)
- 2. Reproducibility of computational workflows is automated using continuous analysis**  
Brett K Beaulieu-Jones, Casey S Greene  
*Nature biotechnology* (2017-04) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6103790/>  
DOI: [10.1038/nbt.3780](https://doi.org/10.1038/nbt.3780) · PMID: [28288103](https://pubmed.ncbi.nlm.nih.gov/28288103/) · PMCID: [PMC6103790](https://pubmed.ncbi.nlm.nih.gov/PMC6103790/)
- 3. Bitcoin for the biological literature.**  
Douglas Heaven  
*Nature* (2019-02) <https://www.ncbi.nlm.nih.gov/pubmed/30718888>  
DOI: [10.1038/d41586-019-00447-9](https://doi.org/10.1038/d41586-019-00447-9) · PMID: [30718888](https://pubmed.ncbi.nlm.nih.gov/30718888/)
- 4. Plan S: Accelerating the transition to full and immediate Open Access to scientific publications**  
cOAlition S  
(2018-09-04) <https://www.wikidata.org/wiki/Q56458321>
- 5. Open access**  
Peter Suber  
*MIT Press* (2012)  
ISBN: 9780262517638
- 6. Open collaborative writing with Manubot**  
Daniel S Himmelstein, Vincent Rubinetti, David R Slochower, Dongbo Hu, Venkat S Malladi, Casey S Greene, Anthony Gitter  
*Manubot* (2020-05-25) <https://greenelab.github.io/meta-review/>
- 7. Opportunities and obstacles for deep learning in biology and medicine**  
Travers Ching, Daniel S Himmelstein, Brett K Beaulieu-Jones, Alexandr A Kalinin, Brian T Do, Gregory P Way, Enrico Ferrero, Paul-Michael Agapow, Michael Zietz, Michael M Hoffman, ... Casey S Greene  
*Journal of The Royal Society Interface* (2018-04) <https://doi.org/gddkhn>  
DOI: [10.1098/rsif.2017.0387](https://doi.org/10.1098/rsif.2017.0387) · PMID: [29618526](https://pubmed.ncbi.nlm.nih.gov/29618526/) · PMCID: [PMC5938574](https://pubmed.ncbi.nlm.nih.gov/PMC5938574/)
- 8. Open collaborative writing with Manubot**  
Daniel S Himmelstein, Vincent Rubinetti, David R Slochower, Dongbo Hu, Venkat S Malladi, Casey S Greene, Anthony Gitter

