Stencila

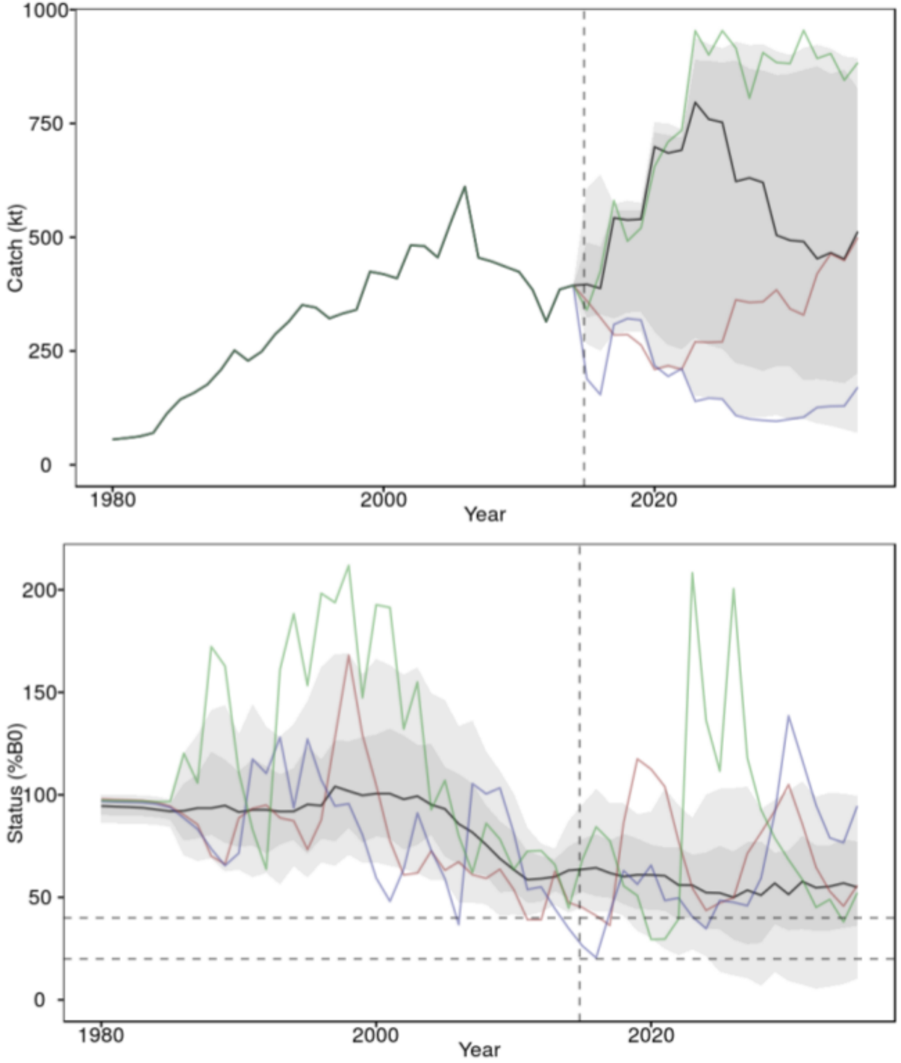
Funded by

ALFRED P. SLOAN FOUNDATION 1934

Alfred P. Sloan
FOUNDATION

https://stenci.la

@stencila

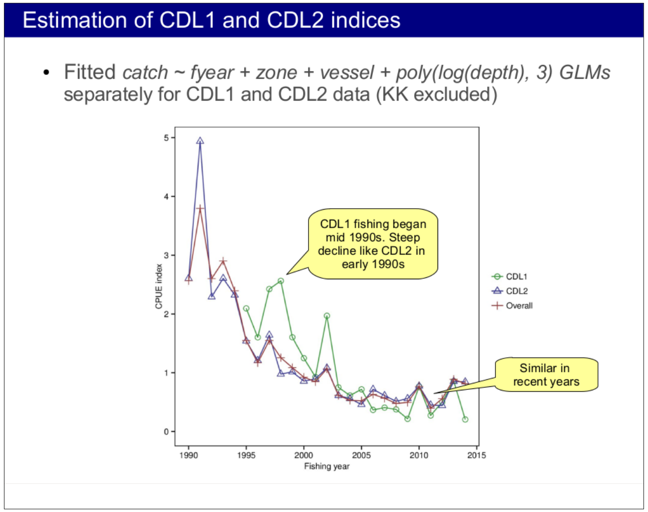

1000
750
500
250
0
Catch (kt)
1980
2000
2020
Year
200
150
100
50
0
Status (%B0)
1980
2000
2020
Year

Context: trying to find a more **efficient,** more **robust** and **transparent** workflow for data-intensive research

Presentations

Project report (150p)

Journal article (12p)



Estimation of CDL1 and CDL2 indices

- Fitted *catch ~ fyear + zone + vessel + poly(log(depth), 3) GLMs* separately for CDL1 and CDL2 data (KK excluded)

CDL1 fishing began mid 1990s. Steep decline like CDL2 in early 1990s

Similar in recent years

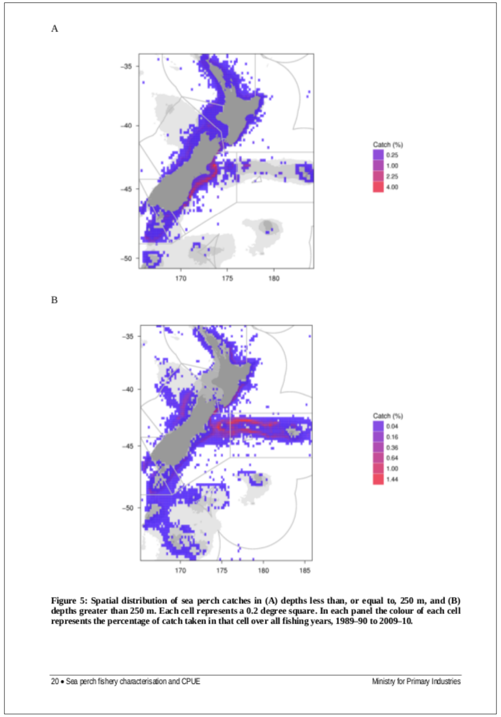
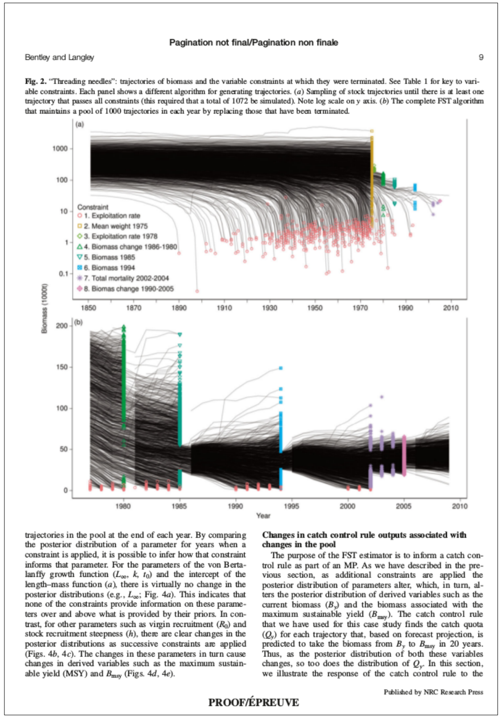


Figure 5: Spatial distribution of sea perch catches in (A) depths less than, or equal to, 250 m, and (B) depths greater than 250 m. Each cell represents a 0.2 degree square. In each panel the colour of each cell represents the percentage of catch taken in that cell over all fishing years, 1989–90 to 2009–10.

20 • Sea perch fishery characterisation and CPUE

Ministry for Primary Industries



Pagination not final/Pagination non finale

Bentley and Langley 9

Fig. 2. "Threading needles": trajectories of biomass and the variable constraints at which they were terminated. See Table 1 for key to variable constraints. Each panel shows a different algorithm for generating trajectories. (*a*) Sampling of stock trajectories until there is at least one trajectory that passes all constraints (this required that a total of 1072 be simulated). Note log scale on *y* axis. (*b*) The complete FST algorithm that maintains a pool of 1000 trajectories in each year by replacing those that have been terminated.

trajectories in the pool at the end of each year. By comparing the posterior distribution of a parameter for years when a constraint is applied, it is possible to infer how that constraint informs that parameter. For the parameters of the von Bertalanffy growth function (L_∞, k, t_0) and the intercept of the length–mass function (a), there is virtually no change in the posterior distributions (e.g., L_∞; Fig. 4*a*). This indicates that none of the constraints provide information on these parameters over and above what is provided by their priors. In contrast, for other parameters such as virgin recruitment (R_0) and stock recruitment steepness (h), there are clear changes in the posterior distributions as successive constraints are applied (Figs. 4*b*, 4*c*). The changes in these parameters in turn cause changes in derived variables such as the maximum sustainable yield (MSY) and B_{msy} (Figs. 4*d*, 4*e*).

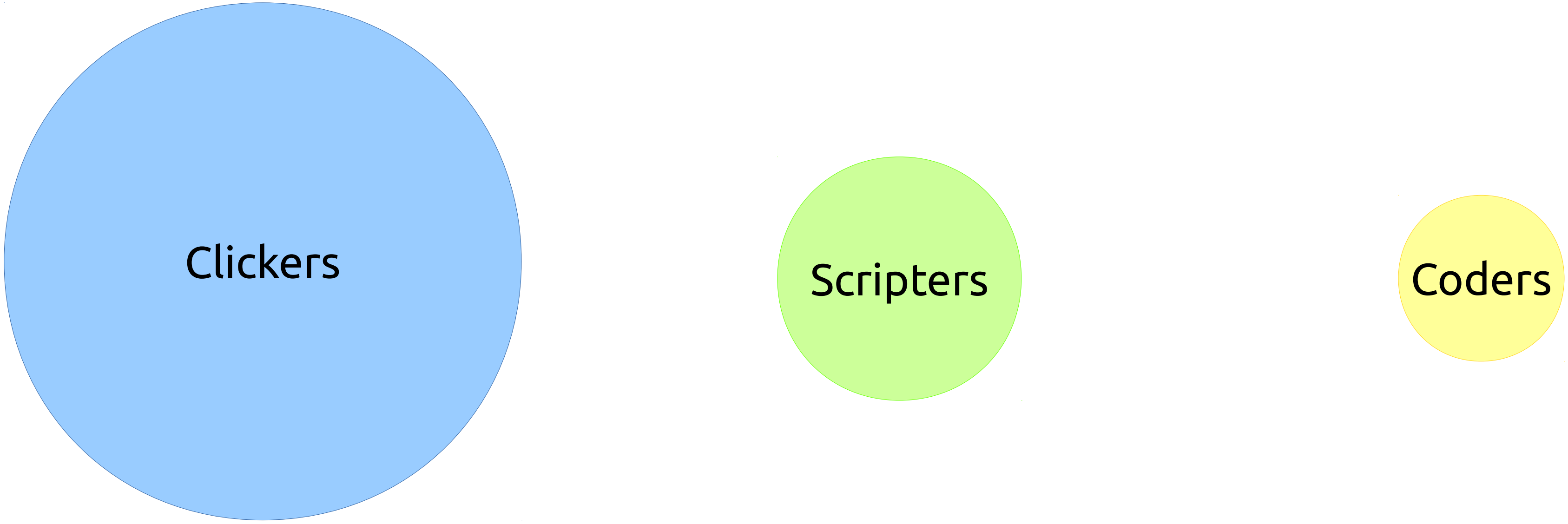
Changes in catch control rule outputs associated with changes in the pool

The purpose of the FST estimator is to inform a catch control rule as part of an MP. As we have described in the previous section, as additional constraints are applied the posterior distribution of parameters alter, which, in turn, alters the posterior distribution of derived variables such as the current biomass (B_y) and the biomass associated with the maximum sustainable yield (B_{msy}). The catch control rule that we have used for this case study finds the catch quota (Q_y) for each trajectory that, based on forecast projection, is predicted to take the biomass from B_y to B_{msy} in 20 years. Thus, as the posterior distribution of both these variables changes, so too does the distribution of Q_y. In this section, we illustrate the response of the catch control rule to the

Published by NRC Research Press

PROOF/ÉPREUVE

Context: different people need different interfaces



Clickers

Scripters

Coders

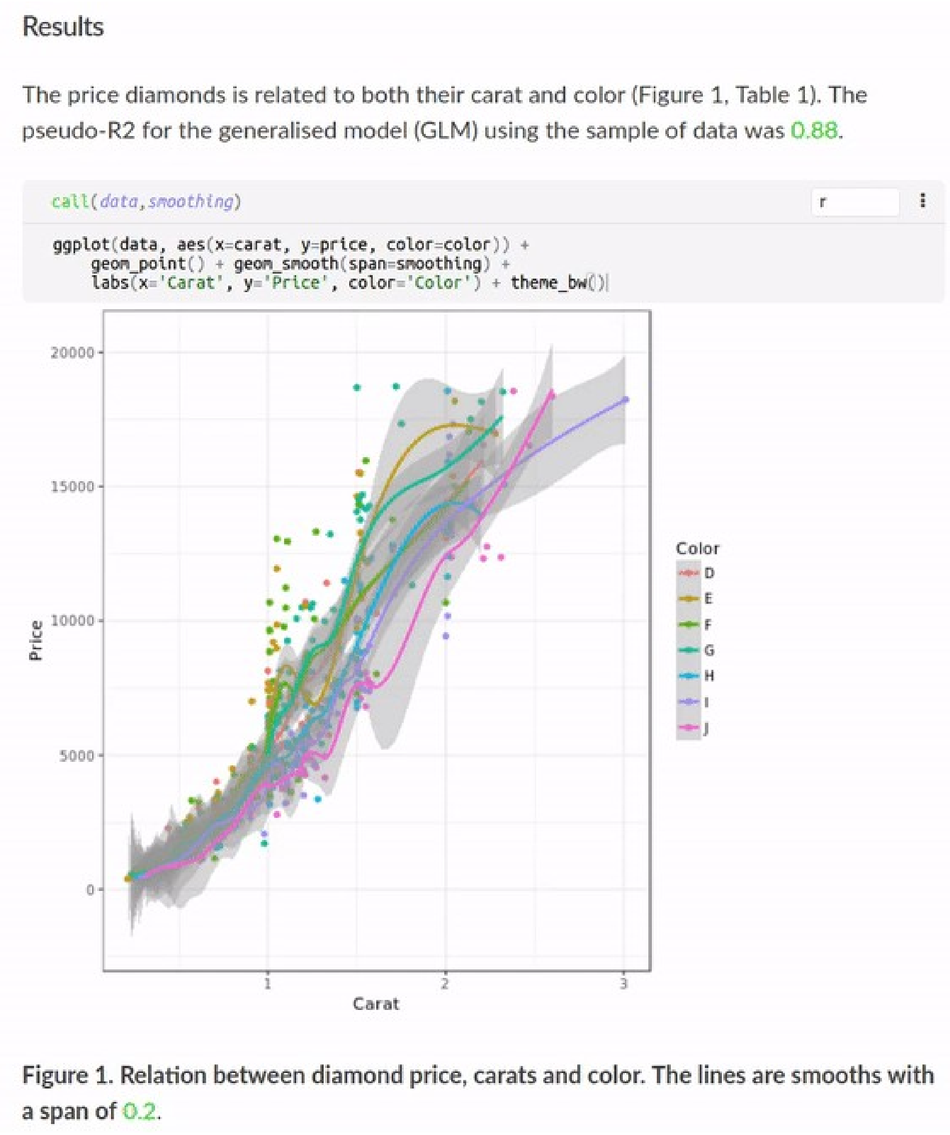
Visual interfaces
Excel, Word

Code + visual interfaces
R scripts + Word

Code interfaces
Rmarkdown, Latex, Git

Little or no reproducibility → High reproducibility

Need to **reduce barriers** to reproducibility and **bridge the gap** between clickers and coders



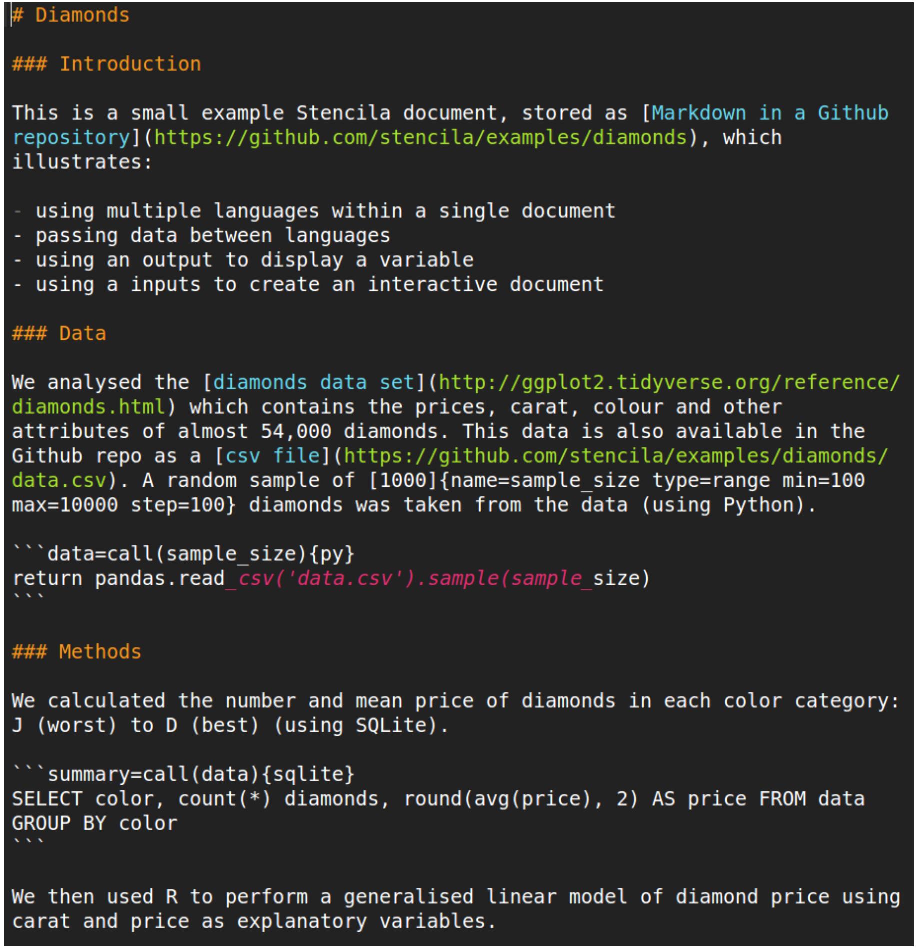
Results

The price diamonds is related to both their carat and color (Figure 1, Table 1). The pseudo-R2 for the generalised model (GLM) using the sample of data was 0.88.

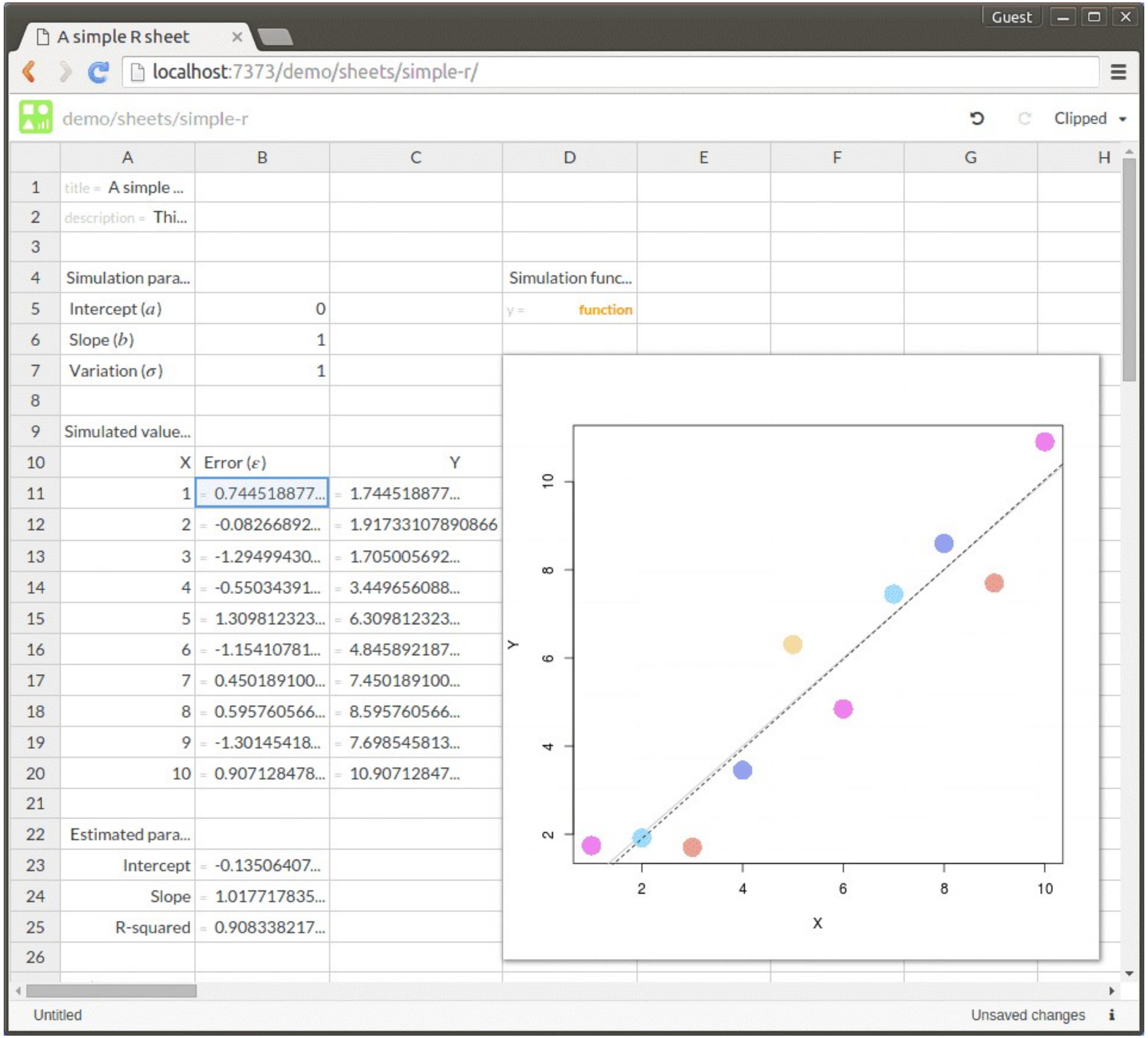
```
call(data,smoothing)
```

```
ggplot(data, aes(x=carat, y=price, color=color)) +
    geom_point() + geom_smooth(span=smoothing) +
    labs(x='Carat', y='Price', color='Color') + theme_bw()
```

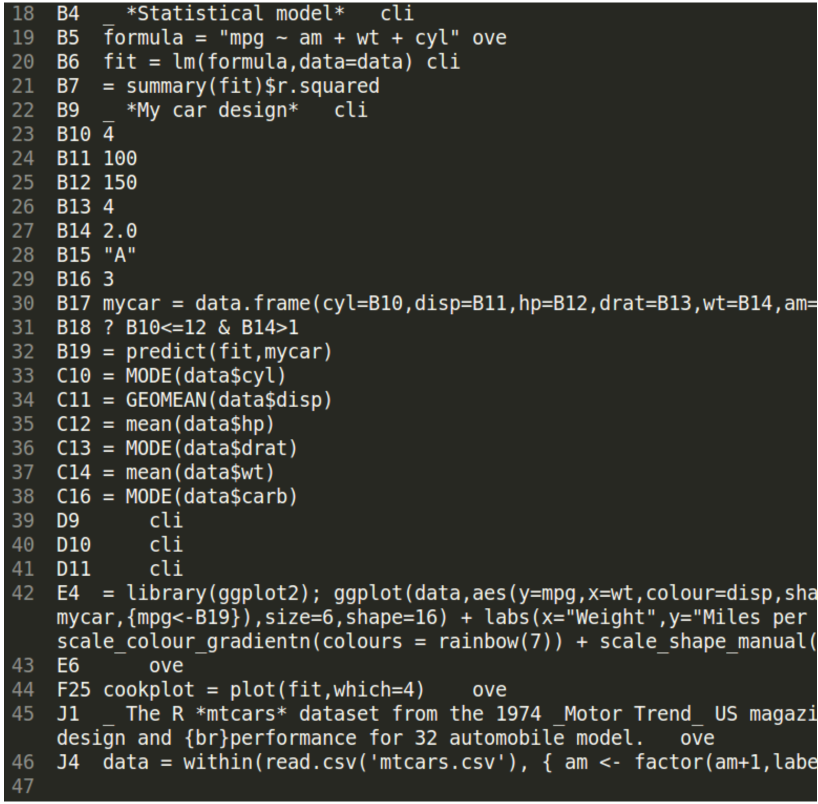
Figure 1. Relation between diamond price, carats and color. The lines are smooths with a span of 0.2.



~~~
# Diamonds

### Introduction

This is a small example Stencila document, stored as [Markdown in a Github
repository](https://github.com/stencila/examples/diamonds), which
illustrates:

- using multiple languages within a single document
- passing data between languages
- using an output to display a variable
- using a inputs to create an interactive document

### Data

We analysed the [diamonds data set](http://ggplot2.tidyverse.org/reference/
diamonds.html) which contains the prices, carat, colour and other
attributes of almost 54,000 diamonds. This data is also available in the
Github repo as a [csv file](https://github.com/stencila/examples/diamonds/
data.csv). A random sample of [1000]{name=sample_size type=range min=100
max=10000 step=100} diamonds was taken from the data (using Python).

```data=call(sample_size){py}
return pandas.read_csv('data.csv').sample(sample_size)
```

### Methods

We calculated the number and mean price of diamonds in each color category:
J (worst) to D (best) (using SQLite).

```summary=call(data){sqlite}
SELECT color, count(*) diamonds, round(avg(price), 2) AS price FROM data
GROUP BY color
```

We then used R to perform a generalised linear model of diamond price using
carat and price as explanatory variables.
~~~

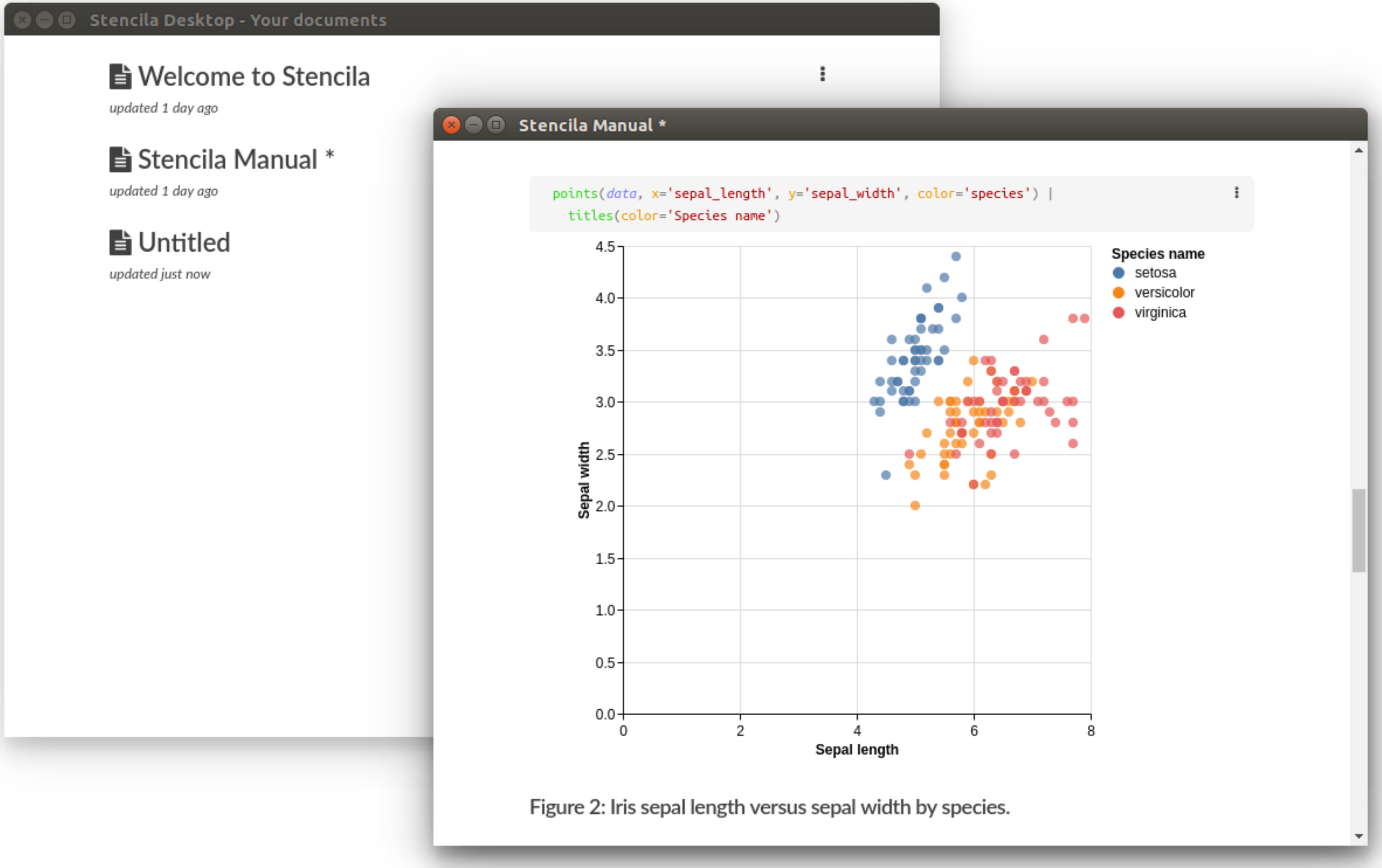


	A	B	C	D	E	F	G	H
1	title = A simple ...							
2	description = Thi...							
3								
4	Simulation para...			Simulation func...				
5	Intercept (a)	0		y = function				
6	Slope (b)	1						
7	Variation (σ)	1						
8								
9	Simulated value...							
10	X	Error (ε)	Y					
11	1	= 0.744518877...	= 1.744518877...					
12	2	= -0.08266892...	= 1.91733107890866					
13	3	= -1.29499430...	= 1.705005692...					
14	4	= -0.55034391...	= 3.449656088...					
15	5	= 1.309812323...	= 6.309812323...					
16	6	= -1.15410781...	= 4.845892187...					
17	7	= 0.450189100...	= 7.450189100...					
18	8	= 0.595760566...	= 8.595760566...					
19	9	= -1.30145418...	= 7.698545813...					
20	10	= 0.907128478...	= 10.90712847...					
21								
22	Estimated para...							
23	Intercept	= -0.13506407...						
24	Slope	= 1.017717835...						
25	R-squared	= 0.908338217...						
26								



```
B4   _ *Statistical model*   cli
B5   formula = "mpg ~ am + wt + cyl" ove
B6   fit = lm(formula,data=data) cli
B7   = summary(fit)$r.squared
B9   _ *My car design*   cli
B10 4
B11 100
B12 150
B13 4
B14 2.0
B15 "A"
B16 3
B17 mycar = data.frame(cyl=B10,disp=B11,hp=B12,drat=B13,wt=B14,am=
B18 ? B10<=12 & B14>1
B19 = predict(fit,mycar)
C10 = MODE(data$cyl)
C11 = GEOMEAN(data$disp)
C12 = mean(data$hp)
C13 = MODE(data$drat)
C14 = mean(data$wt)
C16 = MODE(data$carb)
D9      cli
D10     cli
D11     cli
E4   = library(ggplot2); ggplot(data,aes(y=mpg,x=wt,colour=disp,sha
mycar,{mpg<-B19}),size=6,shape=16) + labs(x="Weight",y="Miles per
scale_colour_gradientn(colours = rainbow(7)) + scale_shape_manual(
E6      ove
F25 cookplot = plot(fit,which=4)    ove
J1   _ The R *mtcars* dataset from the 1974 _Motor Trend_ US magazi
design and {br}performance for 32 automobile model.   ove
J4   data = within(read.csv('mtcars.csv'), { am <- factor(am+1,labe

```

https://github.com/stencila/desktop



Stencila Desktop - Your documents

Welcome to Stencila

updated 1 day ago

Stencila Manual *

updated 1 day ago

Untitled

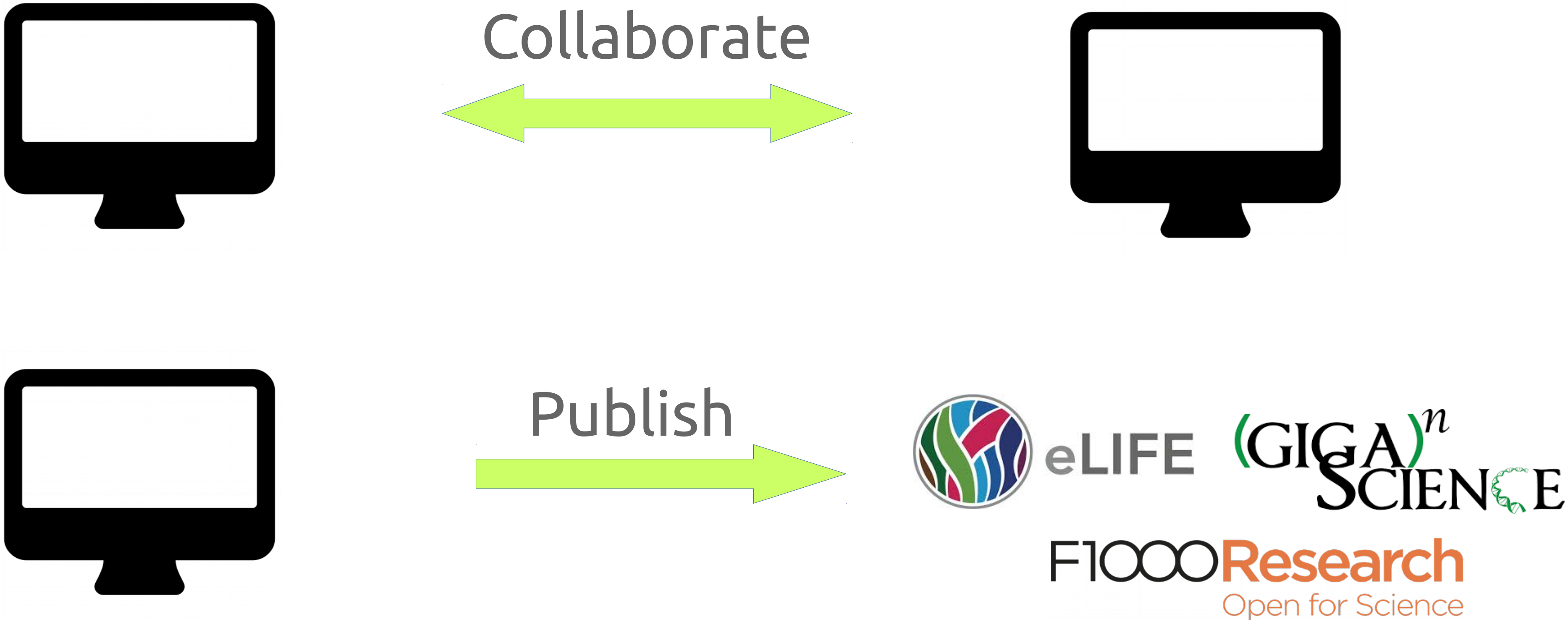
updated just now

Stencila Manual *

```
points(data, x='sepal_length', y='sepal_width', color='species') |
  titles(color='Species name')
```

Figure 2: Iris sepal length versus sepal width by species.

“Okay…, so you have a nice desktop app for reproducible research but how do you reliably…



Collaborate

Publish

eLIFE

(GIGA)^n SCIENCE

F1000Research

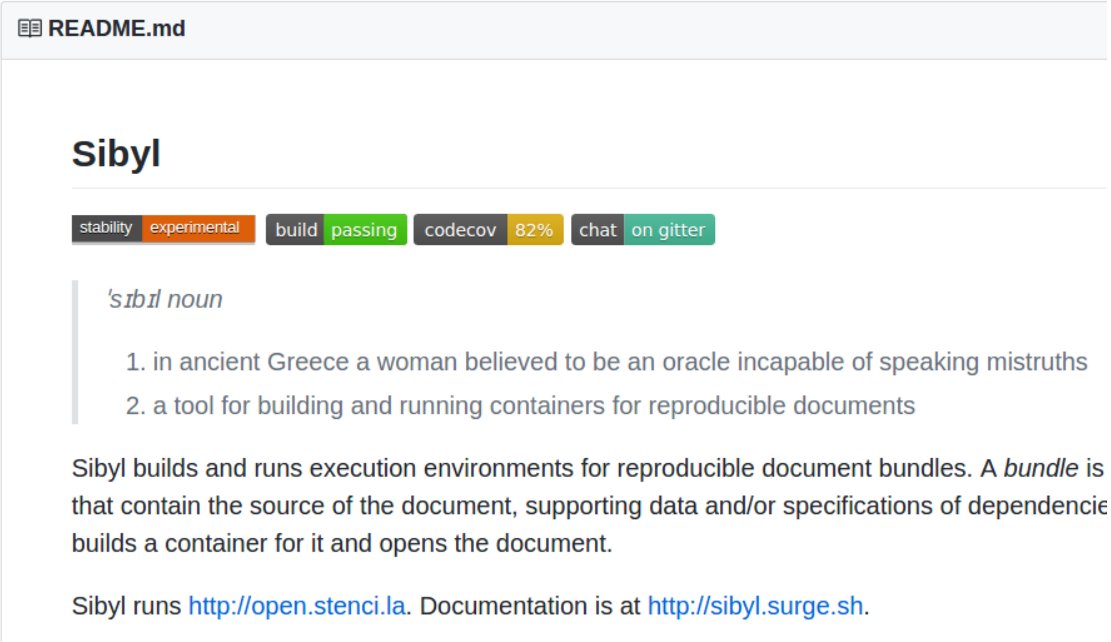
Open for Science

https://github.com/stencila/sibyl

dat DATA

GitHub

Dropbox



📖 README.md

Sibyl

stability experimental | build passing | codecov 82% | chat on gitter

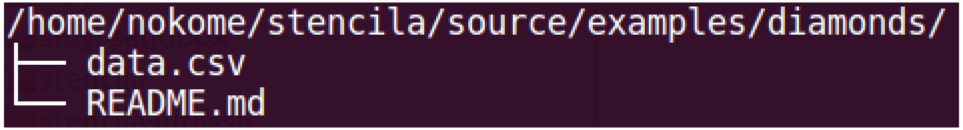
> *ˈsɪbɪl noun*
>
> 1. in ancient Greece a woman believed to be an oracle incapable of speaking mistruths
> 2. a tool for building and running containers for reproducible documents

Sibyl builds and runs execution environments for reproducible document bundles. A *bundle* is that contain the source of the document, supporting data and/or specifications of dependencie builds a container for it and opens the document.

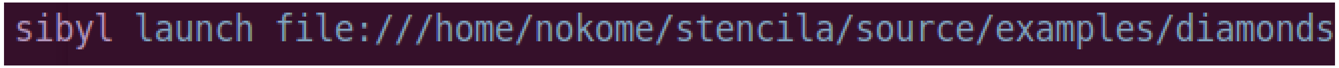
Sibyl runs http://open.stenci.la. Documentation is at http://sibyl.surge.sh.

docker

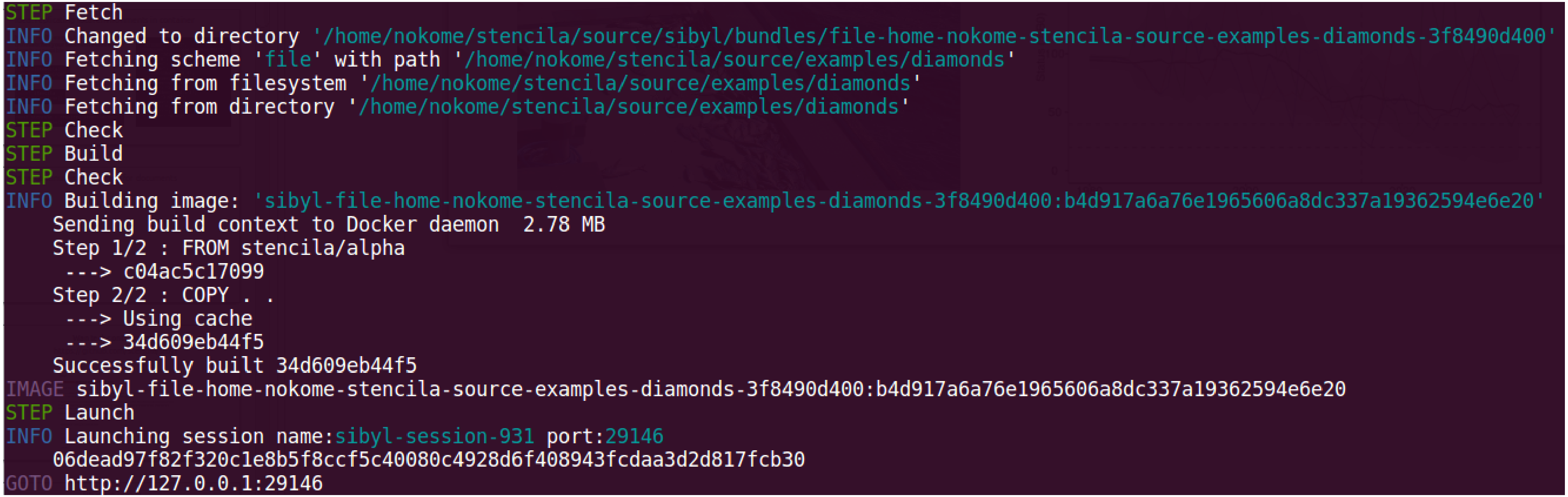
kubernetes



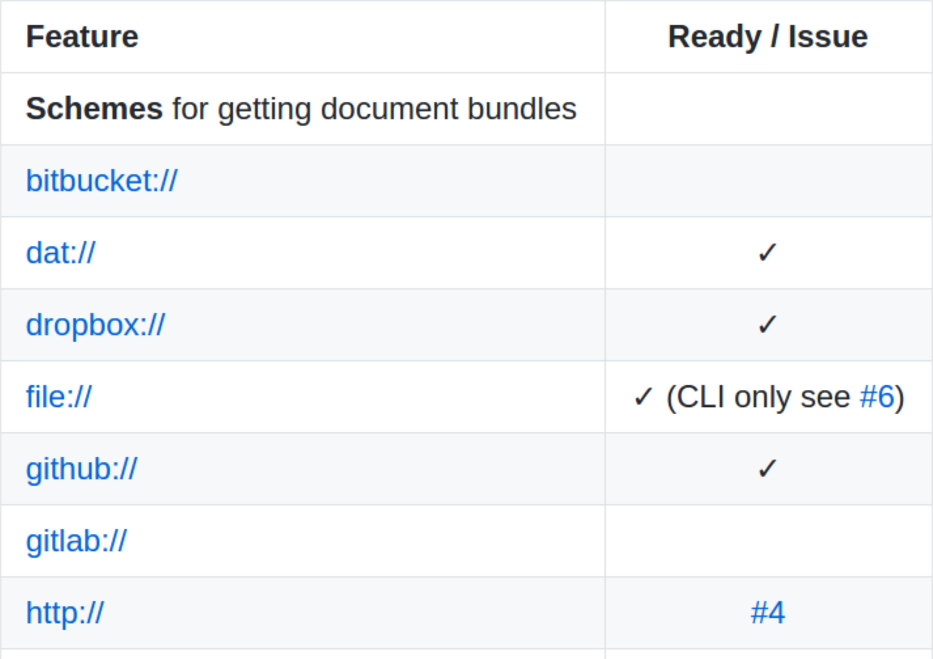
```
/home/nokome/stencila/source/examples/diamonds/
├── data.csv
└── README.md
```



```
sibyl launch file:///home/nokome/stencila/source/examples/diamonds
```



```
STEP Fetch
INFO Changed to directory '/home/nokome/stencila/source/sibyl/bundles/file-home-nokome-stencila-source-examples-diamonds-3f8490d400'
INFO Fetching scheme 'file' with path '/home/nokome/stencila/source/examples/diamonds'
INFO Fetching from filesystem '/home/nokome/stencila/source/examples/diamonds'
INFO Fetching from directory '/home/nokome/stencila/source/examples/diamonds'
STEP Check
STEP Build
STEP Check
INFO Building image: 'sibyl-file-home-nokome-stencila-source-examples-diamonds-3f8490d400:b4d917a6a76e1965606a8dc337a19362594e6e20'
    Sending build context to Docker daemon  2.78 MB
    Step 1/2 : FROM stencila/alpha
     ---> c04ac5c17099
    Step 2/2 : COPY . .
     ---> Using cache
     ---> 34d609eb44f5
    Successfully built 34d609eb44f5
IMAGE sibyl-file-home-nokome-stencila-source-examples-diamonds-3f8490d400:b4d917a6a76e1965606a8dc337a19362594e6e20
STEP Launch
INFO Launching session name:sibyl-session-931 port:29146
    06dead97f82f320c1e8b5f8ccf5c40080c4928d6f408943fcdaa3d2d817fcb30
GOTO http://127.0.0.1:29146
```



Feature	Ready / Issue
Schemes for getting document bundles	
bitbucket://	
dat://	✓
dropbox://	✓
file://	✓ (CLI only see #6)
github://	✓
gitlab://	
http://	#4

Main document resolution:

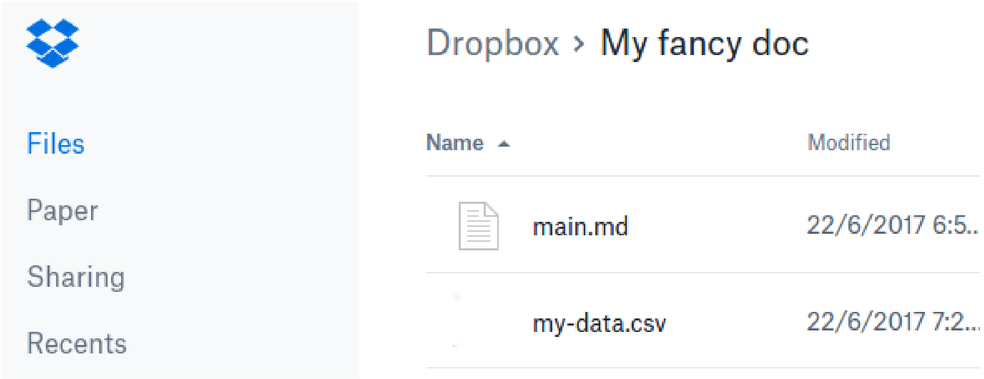
```
main.*
index.*
README.*
```

Main document formats:

```
*.html  *.md
*.Rmd  *.ipynb
```

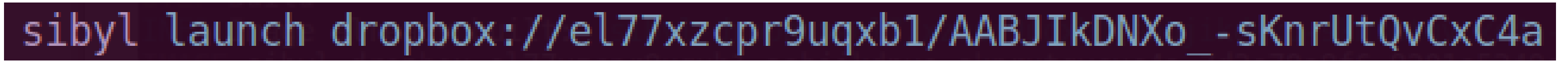
Image customisation:

```
package.json
requirements.txt
r-requires.txt
Dockerfile
```

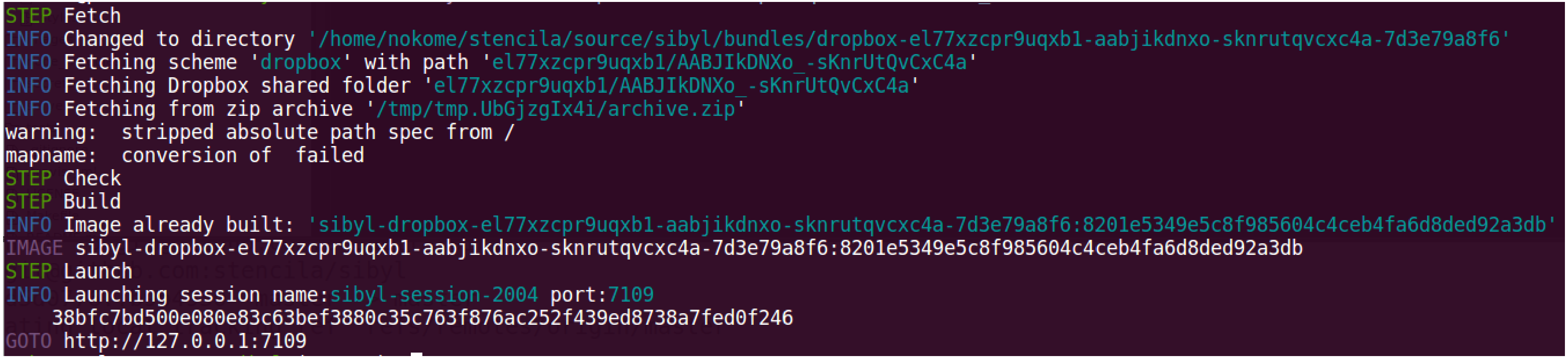


Dropbox › My fancy doc

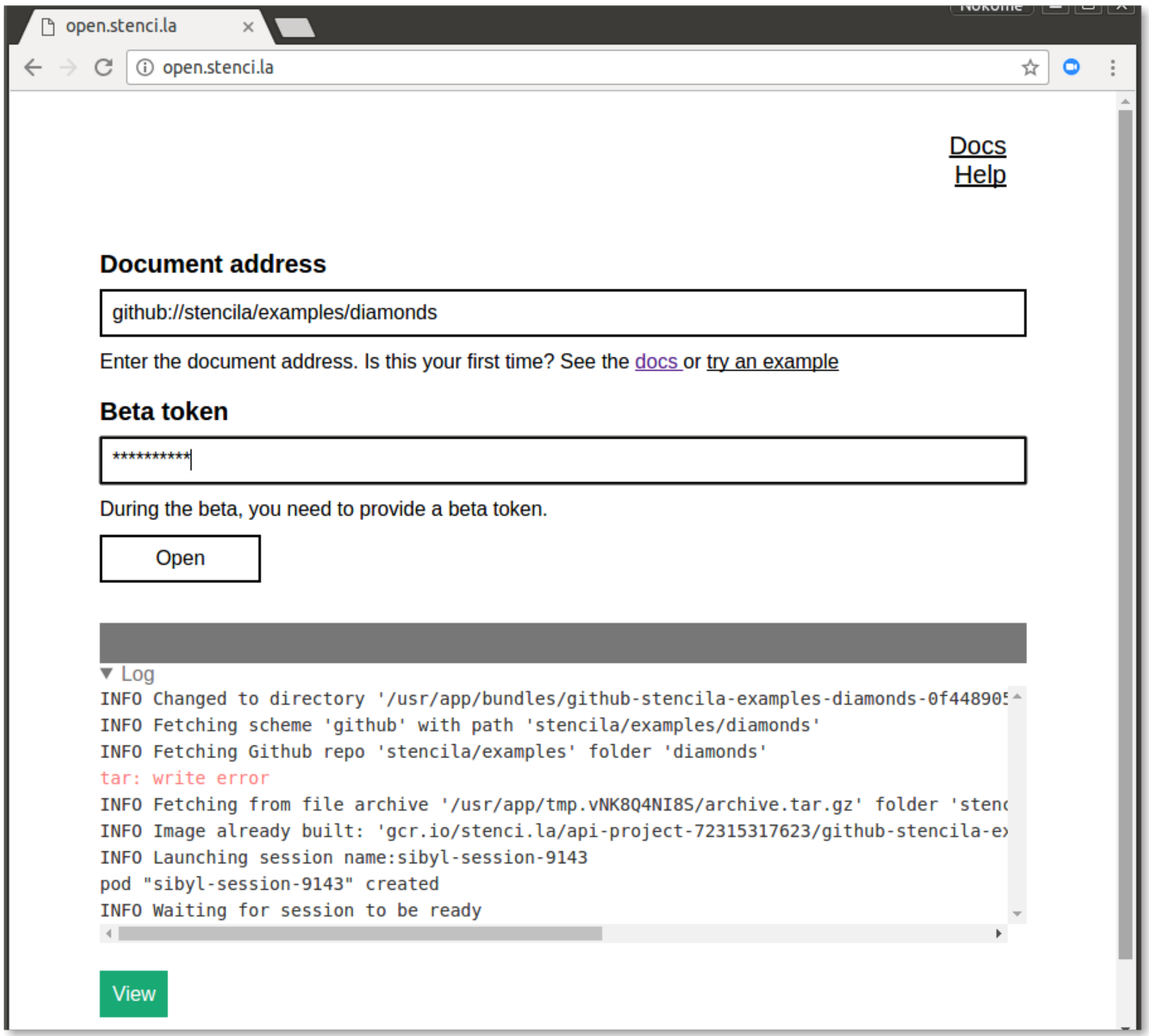
Name ▲	Modified
main.md	22/6/2017 6:5..
my-data.csv	22/6/2017 7:2...



```
sibyl launch dropbox://el77xzcpr9uqxb1/AABJIkDNXo_-sKnrUtQvCxC4a
```



```
STEP Fetch
INFO Changed to directory '/home/nokome/stencila/source/sibyl/bundles/dropbox-el77xzcpr9uqxb1-aabjikdnxo-sknrutqvcxc4a-7d3e79a8f6'
INFO Fetching scheme 'dropbox' with path 'el77xzcpr9uqxb1/AABJIkDNXo_-sKnrUtQvCxC4a'
INFO Fetching Dropbox shared folder 'el77xzcpr9uqxb1/AABJIkDNXo_-sKnrUtQvCxC4a'
INFO Fetching from zip archive '/tmp/tmp.UbGjzgIx4i/archive.zip'
warning:  stripped absolute path spec from /
mapname:  conversion of  failed
STEP Check
STEP Build
INFO Image already built: 'sibyl-dropbox-el77xzcpr9uqxb1-aabjikdnxo-sknrutqvcxc4a-7d3e79a8f6:8201e5349e5c8f985604c4ceb4fa6d8ded92a3db'
IMAGE sibyl-dropbox-el77xzcpr9uqxb1-aabjikdnxo-sknrutqvcxc4a-7d3e79a8f6:8201e5349e5c8f985604c4ceb4fa6d8ded92a3db
STEP Launch
INFO Launching session name:sibyl-session-2004 port:7109
    38bfc7bd500e080e83c63bef3880c35c763f876ac252f439ed8738a7fed0f246
GOTO http://127.0.0.1:7109
```



Document address

github://stencila/examples/diamonds

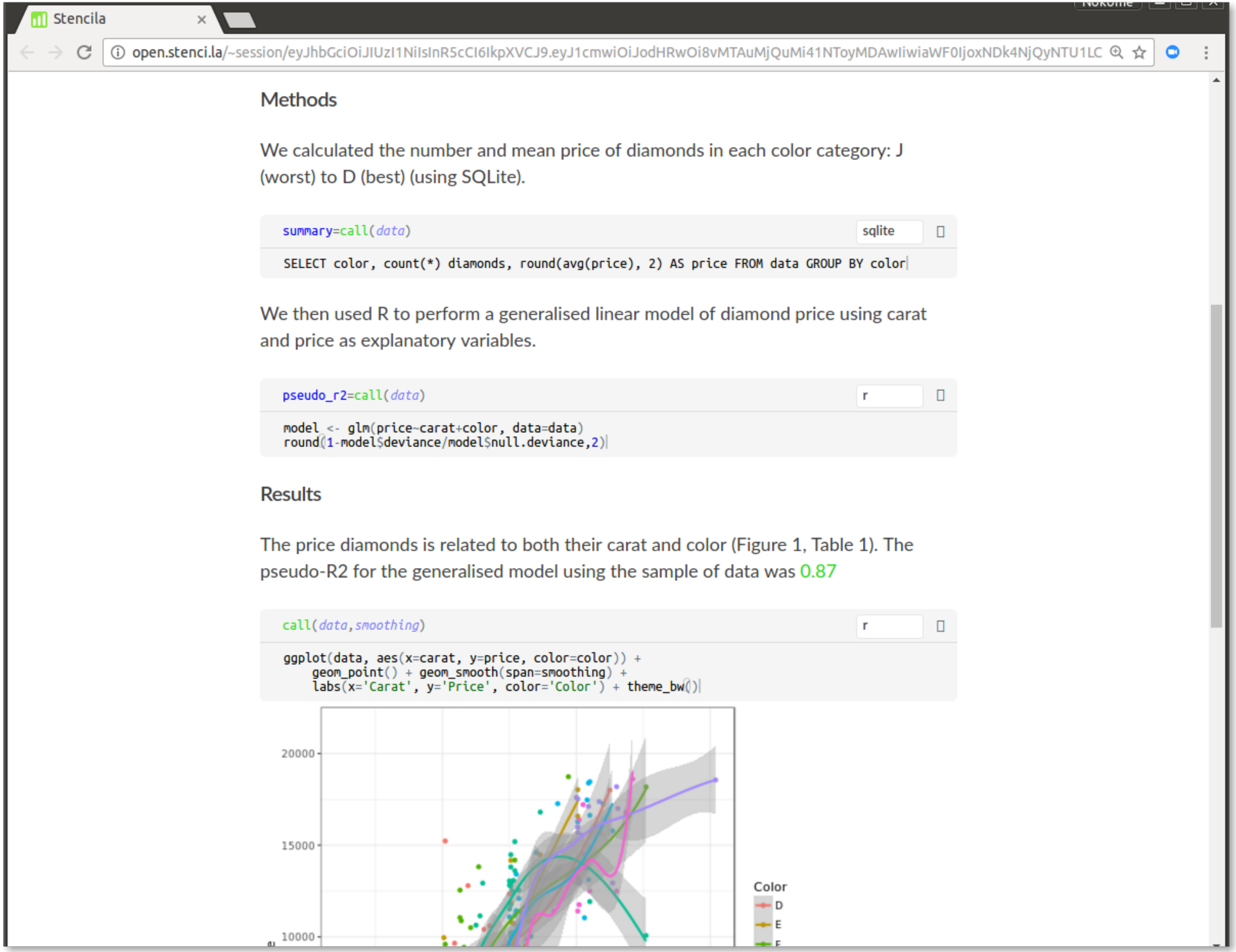
Enter the document address. Is this your first time? See the docs or try an example

Beta token

****

During the beta, you need to provide a beta token.

```
INFO Changed to directory '/usr/app/bundles/github-stencila-examples-diamonds-0f44890
INFO Fetching scheme 'github' with path 'stencila/examples/diamonds'
INFO Fetching Github repo 'stencila/examples' folder 'diamonds'
tar: write error
INFO Fetching from file archive '/usr/app/tmp.vNK8Q4NI8S/archive.tar.gz' folder 'stenc
INFO Image already built: 'gcr.io/stenci.la/api-project-72315317623/github-stencila-ex
INFO Launching session name:sibyl-session-9143
pod "sibyl-session-9143" created
INFO Waiting for session to be ready
```


Stencila
open.stenci.la/~session/eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJ1cmwiOiJodHRwOi8vMTAuMjQuMi41NToyMDAwIiwiaWF0IjoxNDk4NjQyNTU1LC

Methods

We calculated the number and mean price of diamonds in each color category: J (worst) to D (best) (using SQLite).

```
summary=call(data)                                                    sqlite

SELECT color, count(*) diamonds, round(avg(price), 2) AS price FROM data GROUP BY color
```

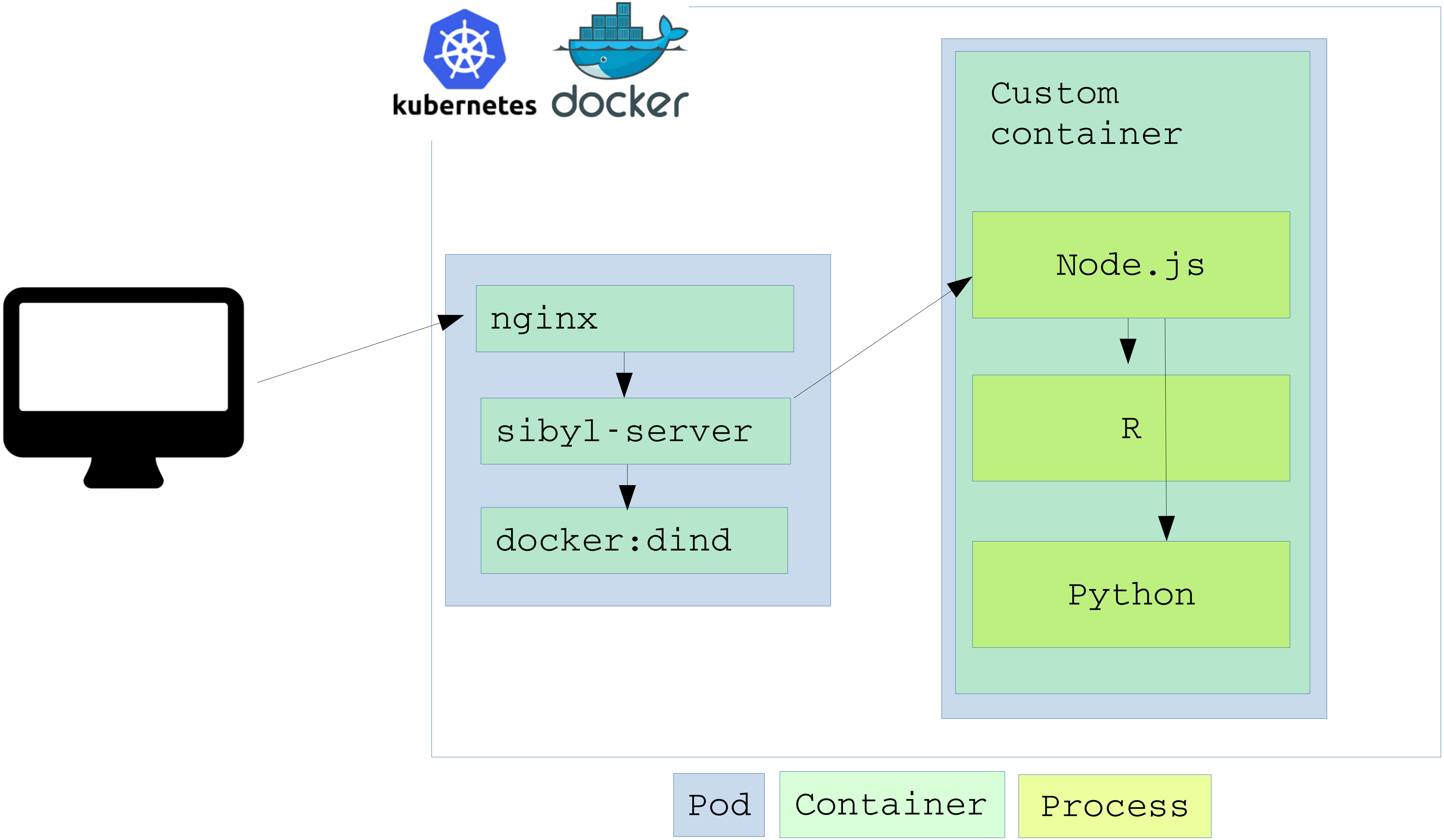
We then used R to perform a generalised linear model of diamond price using carat and price as explanatory variables.

```
pseudo_r2=call(data)                                                  r

model <- glm(price~carat+color, data=data)
round(1-model$deviance/model$null.deviance,2)
```

Results

The price diamonds is related to both their carat and color (Figure 1, Table 1). The pseudo-R2 for the generalised model using the sample of data was 0.87

```
call(data,smoothing)                                                  r

ggplot(data, aes(x=carat, y=price, color=color)) +
    geom_point() + geom_smooth(span=smoothing) +
    labs(x='Carat', y='Price', color='Color') + theme_bw()
```



kubernetes docker

nginx

sibyl-server

docker:dind

Custom container

Node.js

R

Python

Pod

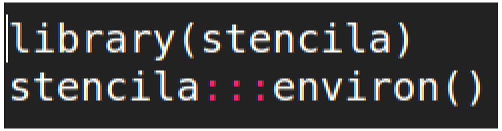
Container

Process

Next steps: replicating local environments in container

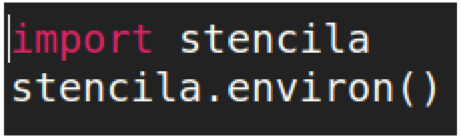
Make it **really easy** to build a container that **matches your local environment** as closely as possible

R



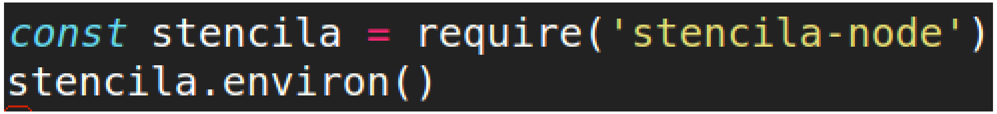
```
library(stencila)
stencila:::environ()
```

Python

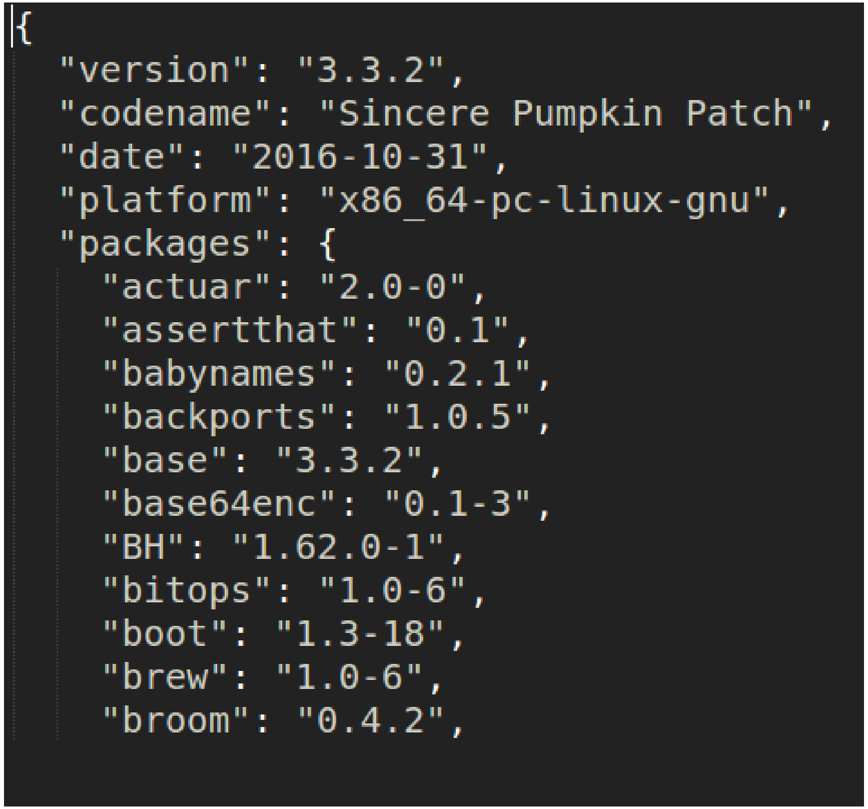


```
import stencila
stencila.environ()
```

Node.js



```
const stencila = require('stencila-node')
stencila.environ()
```



```
{
  "version": "3.3.2",
  "codename": "Sincere Pumpkin Patch",
  "date": "2016-10-31",
  "platform": "x86_64-pc-linux-gnu",
  "packages": {
    "actuar": "2.0-0",
    "assertthat": "0.1",
    "babynames": "0.2.1",
    "backports": "1.0.5",
    "base": "3.3.2",
    "base64enc": "0.1-3",
    "BH": "1.62.0-1",
    "bitops": "1.0-6",
    "boot": "1.3-18",
    "brew": "1.0-6",
    "broom": "0.4.2",
```

Next steps: continuous integration for documents

- Webhooks to trigger builds - "Travis CI for Clickers"
- Test of **reproducibility** (does doc render?)
- Test **assertions** within documents (does doc do what it is meant to?)

Next steps: daily builds of comprehensive images

- Several images that meet the needs of 90% of use cases: possible?
- Daily image builds tagged with date to allow users to **pin to date** with an `image.txt: stencila/delta==2017-06-26`
- Record package versions on each day – help to determine which package change broke your doc

WANTED

Contributors

https://github.com/stencila/sibyl

Beta testers

Talk to me or email me:
nokome@stenci.la