

```
jeroen at strata in ~ $ learn-shell-for-data-science --title
```

# 50 reasons to learn the shell for doing data science

```
jeroen at strata in ~ $ learn-shell-for-data-science --speaker
```

**Jeroen Janssens**

**@jeroenhjanssens**

**CEO at Data Science Workshops B.V.**

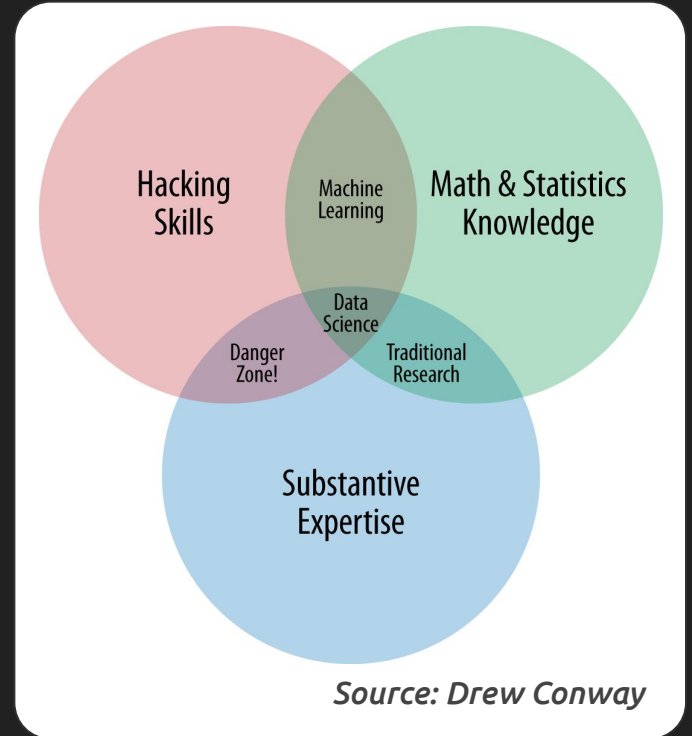
**Author of Data Science at the Command Line**

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 01
```

The shell makes you look  
like a 1337 hacker.

jeroen at strata in ~ \$ learn-shell-for-data-science --reason 02

When it comes to  
**hacking**, the shell  
is indispensable.



```
jeroen at strata in ~ $ learn-shell-for-data-science --osemn
```

## Data science is **OSEMN**:

**O**btaining data

**S**crubbing data

**E**xploring data

**M**odelling data

**i**Nterpreting data

*Source: Mason & Wiggins (2010)*

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 03
```

```
$ pip install scikit-learn
```

```
Requirement already satisfied: scikit-learn in /usr/lib/python3.6/site-packages
```

```
$ cd ~/.ssh
```

```
$ ssh-keygen
```

```
$ cat ~/.ssh/id_rsa.pub | pbcopy
```

```
$ curl -s 'http://api.citybik.es/v2/networks/citi-bike-nyc' |
```

```
> jq '.network.stations[].free_bikes' |
```

```
> paste -sd+ | bc
```

```
7572
```

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 04
```

The shell, with its **read-eval-print-loop**,  
enables you to **play** with your data.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 05
```

The shell is very close to the **filesystem**, which makes it very convenient to work with files on a large scale.



jeroen at strata in ~ \$ learn-shell-for-data-science --reason 06

# Velociraptors.



```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 07
```

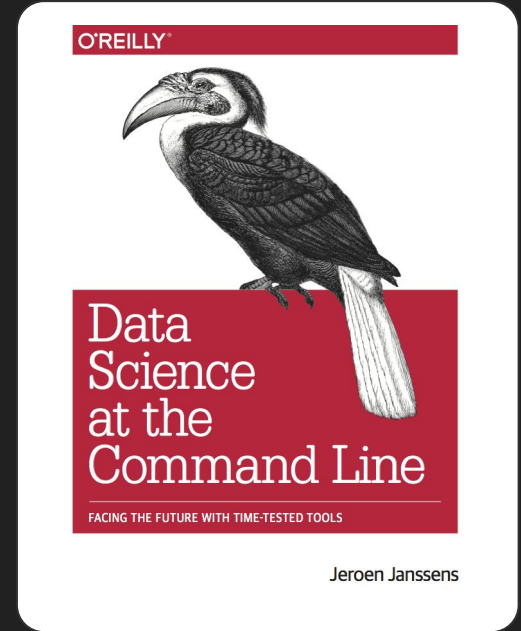
Plenty of great **resources**  
are available to learn the shell.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 08
```

There's a fantastic **book**  
about using the shell  
for doing data science.

Read it for **free** at:

[data science at the command line .com](http://data-science-at-the-command-line.com)



```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 09
```

The shell has a vast  
and interesting **history**.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 10
```

Like **wine**, the shell takes time to be appreciated. Good thing the shell also **ages** like wine.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 11
```

There's always something **new** to learn about the shell and its many tools.

And learning is **fun**.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 12
```

**Docker** containers are great  
for safely learning the shell.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 13
```

The shell gives you access to  
**man** pages, which is like an  
offline **Stack Overflow**.



```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 14
```

[explainshell.com](https://explainshell.com) explains a given command line by matching each argument to the relevant help text in the **man** page.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 15
```

The shell is free.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 16
```

The shell doesn't care whether a tool has been implemented in **Bash**, **C**, **Go**, **Java**, **JavaScript**, **Lisp**, **Perl**, **Python**, **R**, **Rust**, or **Scala**.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 17
```

You can  
customize  
the **hell**  
out of  
the shell.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 18
```

The shell uses **text** as the universal interface, which enables tools from all over the world to **work together** and solve problems.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 19
```

**Most command-line tools do one thing and do it well.**

**The shell is there to let these tools work together in various ways.**

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 20
```

The shell **never bothers you**  
about software updates.

Unless you want it to.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 21
```

The shell gives you great  
**control** over your system.



When **shit** hits the fan  
with **git**, the shell is  
the only interface  
that can clean up  
the mess.



```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 23
```

You can also **program** in the shell.  
A simple **for**-loop can do miracles.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 24
```

Want to **parallelize** or **distribute** your task to multiple cores or machines? Use the shell with a pinch of **parallel**.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 25
```

The shell: come for the **tools**,  
stay for the **environment**.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 26
```

By **default**, the shell comes with many great tools such as **find**, **grep**, and **cut**.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 27
```

Package managers such as **apt-get**,  
**brew**, and **pacman** make it a pleasure to  
install **additional** command-line tools.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 28
```

**New tools** are being developed  
every day for the shell.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 29
```

The shell keeps a **history**.



```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 30
```

You can easily **extend** the shell with your own tools, making you a more **efficient** and **effective** data scientist.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 31
```

The shell lets you **quickly** find out things like: the size of a directory, the encoding of a CSV file, and the resolution of an image.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 32
```

The shell lets you query **databases**,  
access **APIs**, open **remote sheets**, and  
even **scrape** websites.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 33
```

With tools like

`csvkit`, `jq`, and `xmlstarlet`,

you can easily wrangle

`CSV`, `JSON`, and `XML` in the shell.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 34
```

**csvsql** allows you to perform  
**SQL** queries directly on  
**CSV** files in the shell.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 35
```

`telnet towel.blinkenlights.nl`

lets you watch **Star Wars IV**.

Use the shell, Luke.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 36
```

The shell isn't just **available** on UNIX machines and supercomputers. It can also be found on **macOS**, **Raspberry Pi**, and even **Windows 10**.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 37
```

Sometimes the shell **outperforms** fancy  
**big data** technologies.



```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 38
```

You can easily invoke **Python** and **R**  
from the shell.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 39
```

Want to continue working in your favourite **programming language** or **statistical environment**? The shell is totally **cool** with that.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 40
```

You can easily invoke the shell from  
**Jupyter Notebook** and **RStudio**.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 41
```

```
$ echo data science at the command line | cowsay
```

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 41
```

```
$ echo data science at the command line | cowsay
```

```
-----  
< data science at the command line >
```

```
-----
```

```
  \      ^  ^  
  \      --  
  \      (oo)\_____
```



```
  \      ( __ )\      )\ /\
```

The ASCII art cow face consists of several lines of characters. The top line has a backslash followed by two carets. The second line has a backslash followed by two carets and a horizontal line. The third line has a backslash followed by two parentheses containing 'oo' and a horizontal line. The fourth line has a backslash followed by two parentheses containing a horizontal line and a horizontal line, followed by a backslash and a forward slash. The fifth line has two vertical bars, a horizontal line with a 'w' in the middle, and a vertical bar. The sixth line has two vertical bars and a vertical bar.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 42
```

These days, many  
frontend developers  
also use the shell.

jeroen at strata in ~ \$ learn-shell-for-data-science --reason 43

Invoke **sudo** and  
the shell will make  
you a **sandwich**.

MAKE ME A SANDWICH.

SUDO MAKE ME  
A SANDWICH.

WHAT? MAKE  
IT YOURSELF.

OKAY.



*Source: XKCD*

*Note: Do not try on frontend developers*

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 44
```

You can **automate** just about everything using the shell.



```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 45
```

Good luck managing a gazillion instances on **AWS**, **Azure**, and **Google Cloud** using the mouse.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 46
```

The shell often requires **less typing**  
than a programming language.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 47
```

The shell allows you to rename  
750 files with just **three lines** of code.

**Or one**, if you have the right tool.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 48
```

Your **wrists** will thank you  
for using the shell.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 49
```

The shell has been around for almost **50 years**, and probably will be around for the **rest of your career**.

```
jeroen at strata in ~ $ learn-shell-for-data-science --reason 50
```

Because  
Tim  
says so.



A screenshot of a tweet from Tim O'Reilly (@timoreilly) on Twitter. The tweet is a reply to @jeroenhjanssens and @OReillyMedia. The text of the tweet reads: "Anyone who does not have the command line at their beck-and-call is really missing something." The tweet was posted at 11:57 AM on January 24, 2018, from San Francisco, CA. It has 4 retweets and 14 likes. The interface shows a "Following" button, a dropdown arrow, and a row of profile pictures for users who interacted with the tweet. At the bottom, there are icons for replies (1), retweets (4), likes (14), and a direct message icon.

 **Tim O'Reilly**   
@timoreilly

Following 

Replying to [@jeroenhjanssens](#) [@OReillyMedia](#)

Anyone who does not have the command line at their beck-and-call is really missing something.

11:57 AM - 24 Jan 2018 from [San Francisco, CA](#)

4 Retweets 14 Likes



 1  4  14 

```
jeroen at strata in ~ $ learn-shell-for-data-science --thank-you
```

**Jeroen Janssens**

[@jeroenhjanssens](#)

**CEO** at Data Science Workshops B.V.

**Author** of Data Science at the Command Line