Practical Predictive Analytics Seminar:

Welcome!

Brian Holland, FSA, MAAA

September 23, 2020

SOCIETY OF ACTUARIES®

SOA Antitrust Compliance Guidelines

Active participation in the Society of Actuaries is an important aspect of membership. While the positive contributions of professional societies and associations are well-recognized and encouraged, association activities are vulnerable to close antitrust scrutiny. By their very nature, associations bring together industry competitors and other market participants.

The United States antitrust laws aim to protect consumers by preserving the free economy and prohibiting anti-competitive business practices; they promote competition. There are both state and federal antitrust laws, although state antitrust laws closely follow federal law. The Sherman Act, is the primary U.S. antitrust law pertaining to association activities. The Sherman Act prohibits every contract, combination or conspiracy that places an unreasonable restraint on trade. There are, however, some activities that are illegal under all circumstances, such as price fixing, market allocation and collusive bidding.

There is no safe harbor under the antitrust law for professional association activities. Therefore, association meeting participants should refrain from discussing any activity that could potentially be construed as having an anti-competitive effect. Discussions relating to product or service pricing, market allocations, membership restrictions, product standardization or other conditions on trade could arguably be perceived as a restraint on trade and may expose the SOA and its members to antitrust enforcement procedures.

While participating in all SOA in person meetings, webinars, teleconferences or side discussions, you should avoid discussing competitively sensitive information with competitors and follow these guidelines:

- **Do not** discuss prices for services or products or anything else that might affect prices
- **Do not** discuss what you or other entities plan to do in a particular geographic or product markets or with particular customers.
- **Do not** speak on behalf of the SOA or any of its committees unless specifically authorized to do so.
- **Do** leave a meeting where any anticompetitive pricing or market allocation discussion occurs.
- **Do** alert SOA staff and/or legal counsel to any concerning discussions
- **Do** consult with legal counsel before raising any matter or making a statement that may involve competitively sensitive information.

Adherence to these guidelines involves not only avoidance of antitrust violations, but avoidance of behavior which might be so construed. These guidelines only provide an overview of prohibited activities. SOA legal counsel reviews meeting agenda and materials as deemed appropriate and any discussion that departs from the formal agenda should be scrutinized carefully. Antitrust compliance is everyone's responsibility; however, please seek legal counsel if you have any questions or concerns.

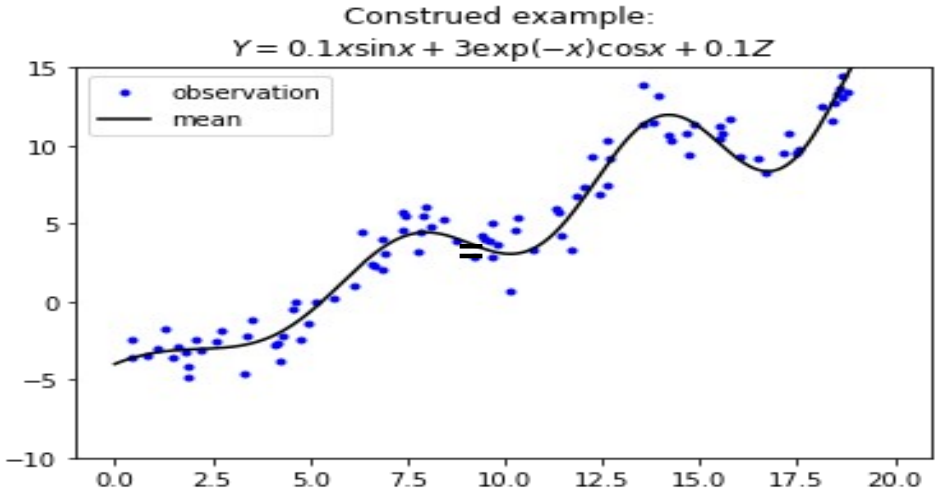
Presentation Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the participants individually and, unless expressly stated to the contrary, are not the opinion or position of the Society of Actuaries, its cosponsors or its committees. The Society of Actuaries does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented. Attendees should note that the sessions are audio-recorded and may be published in various media, including print, audio and video formats without further notice.

Predictive Analytics: Some preliminaries

SOCIETY OF ACTUARIES®

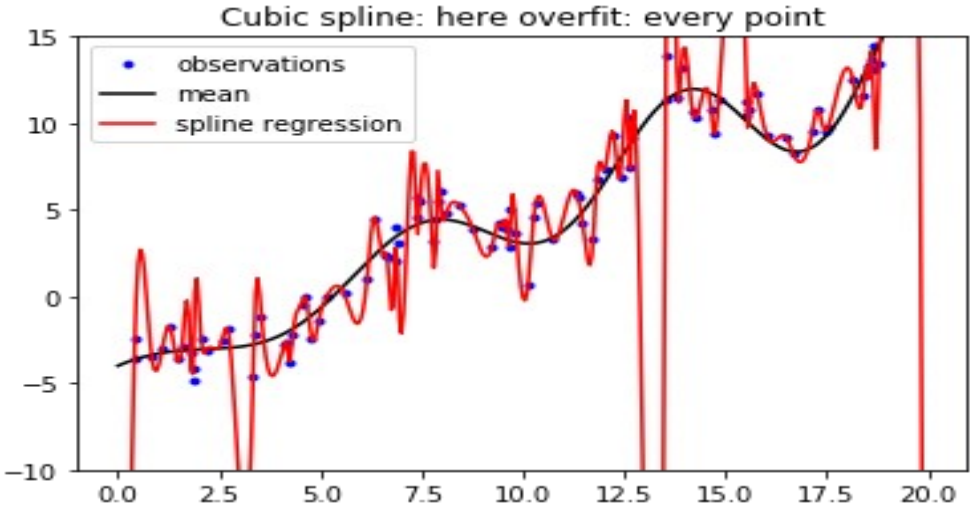
A big topic: underfit vs overfit: How do we balance?



- Polynomial regression
- Spline regression
- Ridge regression
- Tree models
- …

The following slides examine trade-offs with this model.

Let's fit all the points: good job!
… right? *… r i g h t ? … r i g h t ?*

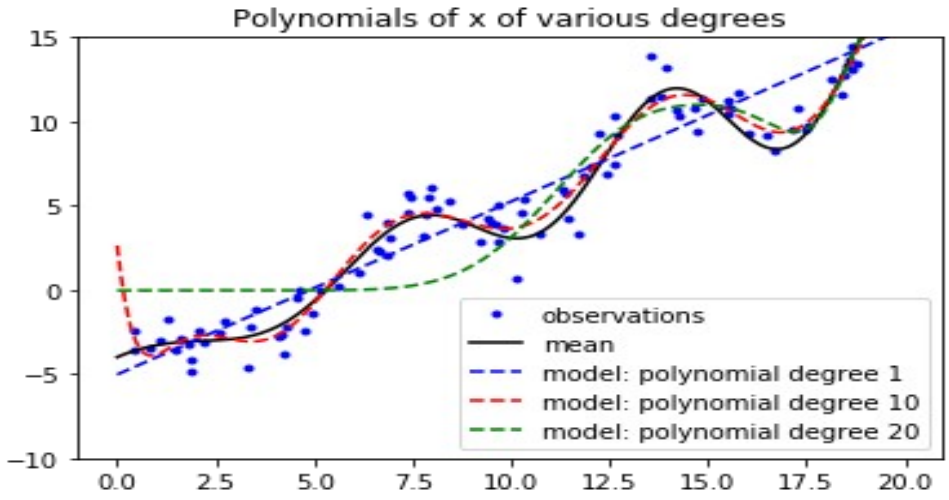


This cubic spline model fits all the points.

Good job? Why, why not?

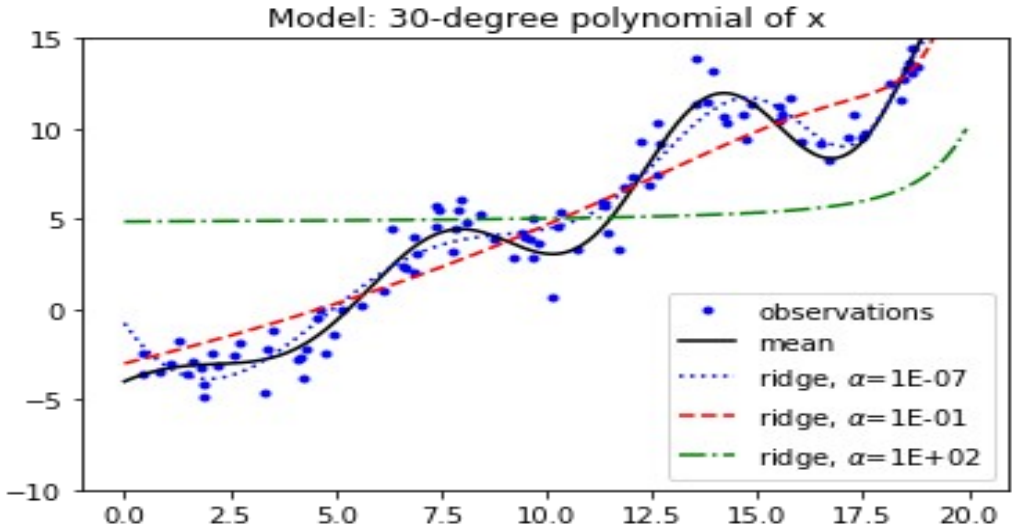
Why not: you don't *really think* that the **new points coming in** would be around the red line.

Simple example: Polynomial Regression



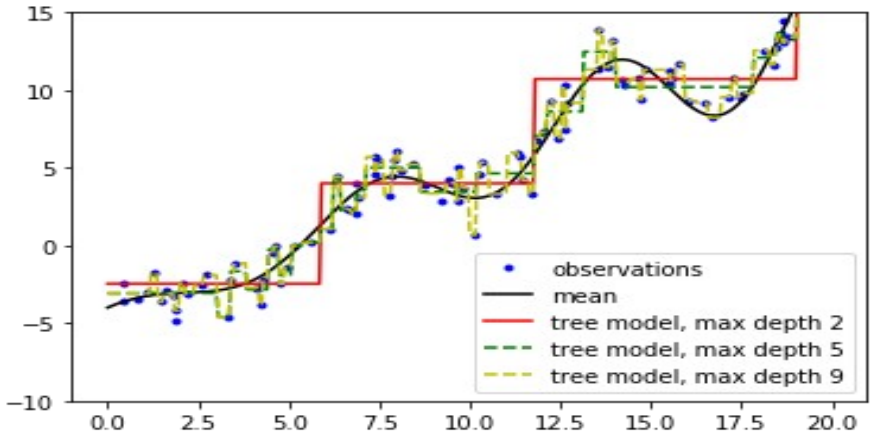
- Higher-degree polynomial terms to capture fit
- Sometimes appear in actuarial work up to second order
- How high is too high?
- Extrapolation – watch out

Simple example: Ridge regression



- More examples of range of options: over- and underfit
- Adding **all** polynomial terms, penalizing coefficients
  - How much penalty is appropriate?
  - Spectrum of choices!
  - Once again – how to choose?

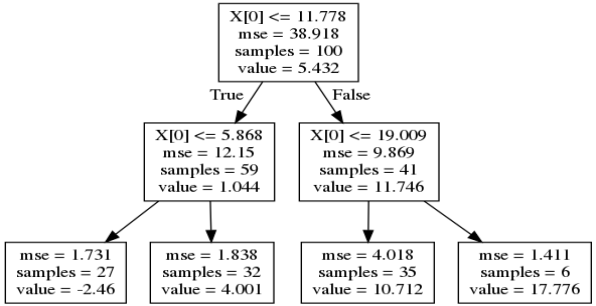
Simple example: Tree

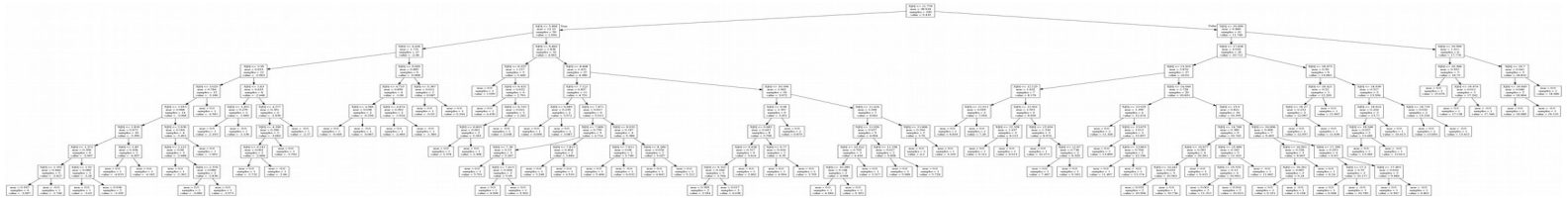


Trade-off in fit – but also in explanation.

Which would you rather explain?

Tree branch could be more than a constant.

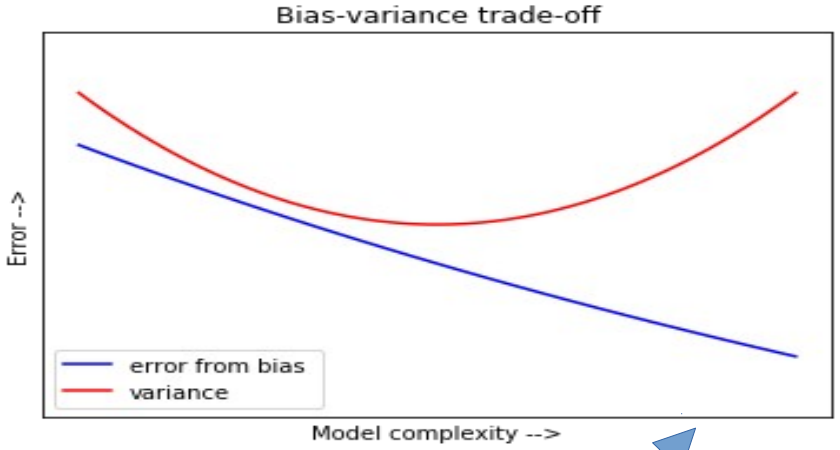


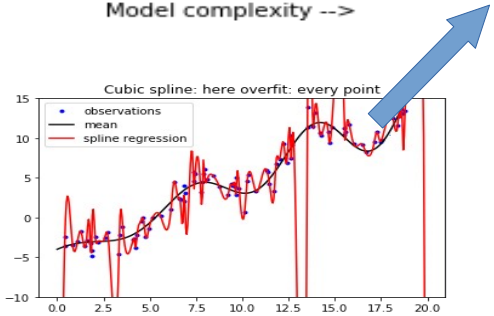


Simple example:
Many others possible

- Other types of models
  - Lasso
  - Random forests
  - Combinations
- Doesn't matter for today – the point is:
  - **The trade-off is there.**
  - **How do you deal with it?**

Framework for addressing trade-off: Bias-Variance Trade-Off





Concept: the trade-off with increasing model complexity **past a point**

Simplest model is a constant: so more complexity should help.

Past a point, more complexity:

- means a better fit to experience

- means worse fit to next year

Bias-Variance Trade-off (cont'd)

In practice:
1. Split data into training and testing sets (3:1, 4:1 or so).
2. Choose complexity to minimize error vs the testing set.

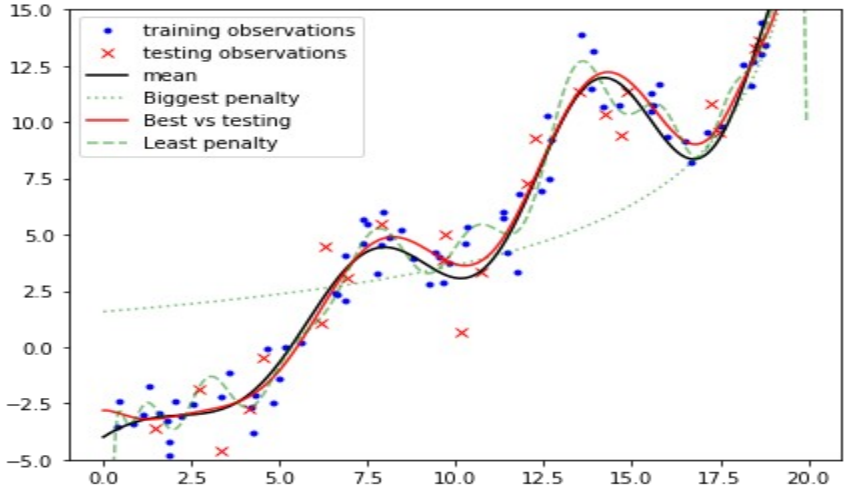
In data-rich scenarios, training, validation, testing sets. *Out of scope today.*

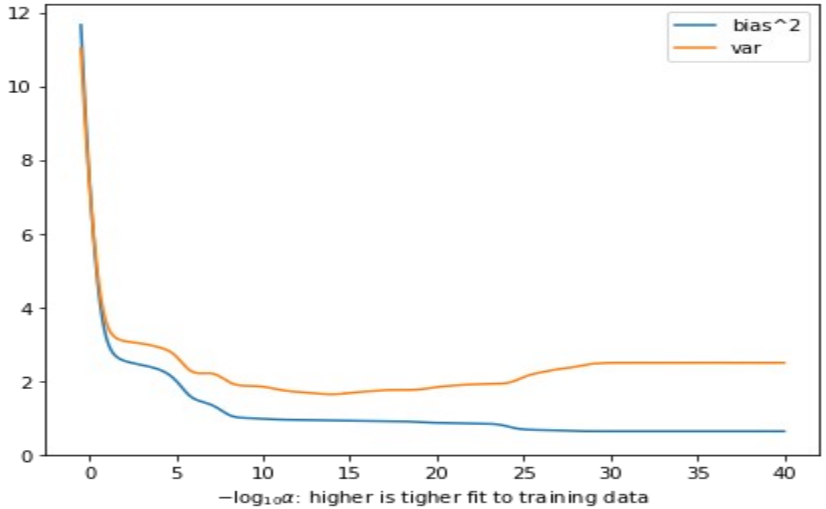
From Elements of Statistical Learning (Hastie et al.):
- **Bias**: mean of the model (a random variable) and underlying reality
- **Variance**: … of the model around the mean of the model

Your model was randomly generated or calibarated in a sense (random split). That might be a talking point.
But remember: it was always random. You're just acknowledging that and dealing with it.

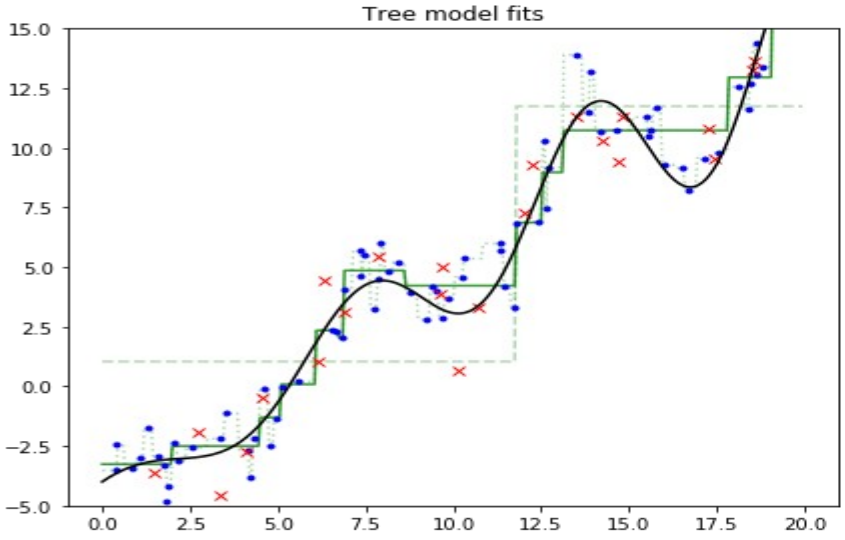
Nice reads
http://scott.fortmann-roe.com/docs/BiasVariance.html
https://jvns.ca/blog/2016/01/02/winning-the-bias-variance-tradeoff/

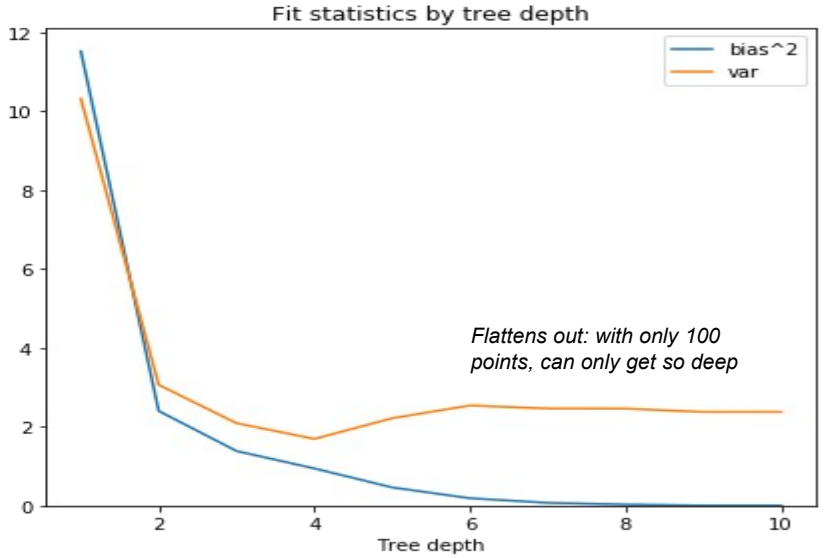
Simple example of framework:
Ridge regression on 30-degree polynomial:
Best fit and worse fits; B-V T-O





Simple example of framework: Tree model





Tools it’s a pity to miss: Git and similar bits

SOCIETY OF ACTUARIES®

SVC

- *Source and Version Control*
- Why use an SVC system: **answer the W's**
- **Who** did **what**, **when**, and **why**
- You document this anyway (right?): the tool helps
- Software packages: discussion and list at… https://en.wikipedia.org/wiki/Version_control
- **Here: focus on *git***

How does this look?

```
* 7e0e1cf Fix README formatting
* bdb0326 (tag: 1.4.0) Bump version to 1.4.0
* cb50131 Adjust readme about YQL errors
*   d044dfc Merge pull request #87 from lukaszbanasiak/yahoo-finance-py35
|\
| * 2703c69 (upstream/yahoo-finance-py35, origin/yahoo-finance-py35) Add support for Python 3.5
* |   1cd8cc6 Merge pull request #86 from lukaszbanasiak/yahoo-finance-20
|\ \
| |/
|/|
| * 57116a4 (upstream/yahoo-finance-20, origin/
|/
*   ec5a4f5 Merge pull request #81 from danielo
|\
| * 4f83eb7 Added new available methods
| * 020d6e4 Removed commented-out lines
| * 02c024e Removed unused data set values avai
| * 0c4d3bf Added additional data set values fr
|/
* 858b9d8 (tag: 1.3.2) version 1.3.2
*   82508ca Merge pull request #76 from lukaszbanasiak/yahoo-finance-75
|\
```

- Many views possible
- Here: my favorite overview
- Shows many branches
- GUI interfaces available

Log: more details

```
commit 57116a47b06b6fa241f7de3d1bc56c95e3a30636
Author: Łukasz Banasiak <lukas.banasiak@gmail.com>
Date:   Thu Nov 17 18:24:33 2016 +0100

    Remove `get_info`

    It's not providing anymore useful information

commit ec5a4f5ed2414488655ab90612cc0c92d3c2eb10
Merge: 858b9d8 4f83eb7
Author: Łukasz Banasiak <lukas.banasiak@gmail.com>
Date:   Thu Nov 17 17:06:46 2016 +0100

    Merge pull request #81 from danielorf/master

    Added additional Yahoo query values

commit 4f83eb773c7cd158ef434ba3d67fc4ff0e5aa718
Author: danielorf <danielorf@gmail.com>
Date:   Sat Nov 5 09:24:42 2016 -0700

    Added new available methods
```

- ID of the change
- Who did it
- When
- Why (comment)
- What: … next slide

What changed?

```
brian@Grinder:~/dev/yahoo-finance$ git diff HEAD^
diff --git a/README.rst b/README.rst
index faa7a8b..899238f 100644
--- a/README.rst
+++ b/README.rst
@@ -128,7 +128,6 @@ Available methods
 - ``get_short_ratio()``
 - ``get_trade_datetime()``
 - ``get_historical(start_date, end_date)``
-- ``get_info()``
 - ``get_name()``
 - ``refresh()``
 - ``get_percent_change_from_year_high()``
brian@Grinder:~/dev/yahoo-finance$
```

This is the change made by the top commit on the prior slide.

It looks like the comment was a good description.

Point: it's easy to check what changed.

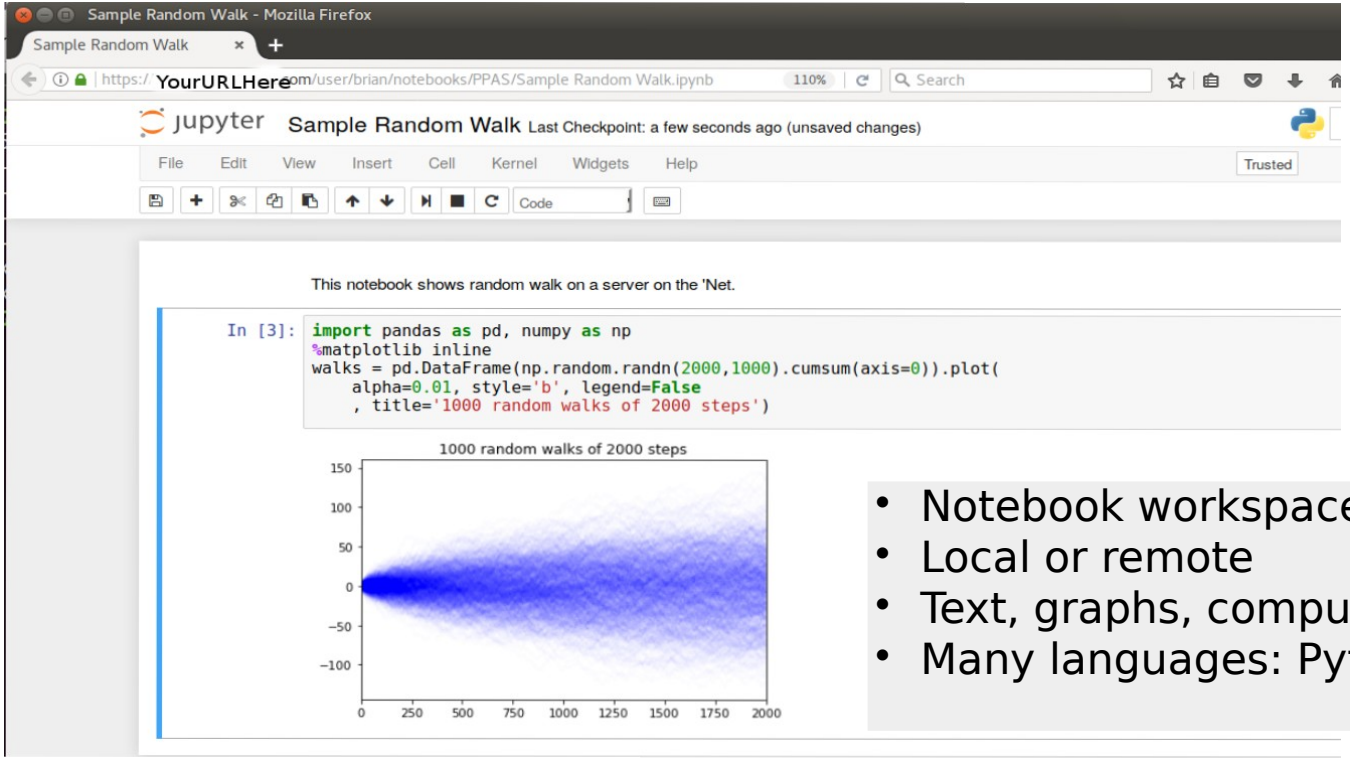
Social networking: github, gitlab

- **Additional features**
- Pull requests
  - to ask others to use your work
  -
- Issue flagging
  - including assigning issues with a due date
- Workflow in general
  - Marking when issues are complete

What do you do with this?

- 1. **Write** your work - text files for source code
- 2. **Commit** your changes – grouped as you like
- - You can "undo" to this point later.
- - You have to comment on the "commit point."
- 3. **Branch** off in a new direction and work on that
- 4. **Check out** a different branch, like “undoing” to another point
- 5. **Merge** other changes into yours.
- 6. **Push** the changes to a place that is shared with others.
- 7. **Pull** others' changes to your local work to synchronize.

Jupyter project



- Notebook workspace
- Local or remote
- Text, graphs, computation all together
- Many languages: Python, R included

Cloud computing

- **Which vendor? Some names:**
  - AWS = Amazon Web Services
  - Microsoft Azure
  - Google, IBM also have offerings
  - Rackspace, DigitalOcean
- **For what**
  - Using a big server for a few hours
  - Having a common workspace with others
  - … many possible reasons: off-site backups, …

Where to learn more

Topic	See...
Git	https://git-scm.com/documentation
Github	https://guides.github.com/activities/hello-world/
Jupyter	https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/ https://jupyterlab.readthedocs.io/en/stable/ (JupyterLab is replacing Jupyter notebooks)
AWS	http://docs.aws.amazon.com/gettingstarted/latest/awsgsg-intro/gsg-aws-intro.html

Thanks for ~~coming~~ attending!

Brian Holland, FSA, MAAA

September 23, 2020

SOCIETY OF ACTUARIES®

SOCIETY OF ACTUARIES®