Dr. Daugherity

Spring 2020

PHYS 453

ACU
ABILENE CHRISTIAN UNIVERSITY

Part 1

WHY ME? WHY PHYSICS?

About Me

Bio:

- Grew up in Oklahoma City
- 2002 ACU: Physics and CS
- 2008 PhD in Nuclear Physics from UT
- Married in 2000, kids are Molly (15) and Ian (10)

Fun Facts:

- Have been paid as a professional model
- Once burned off an eyebrow in class
- Plays drums in Abilene's best cover band

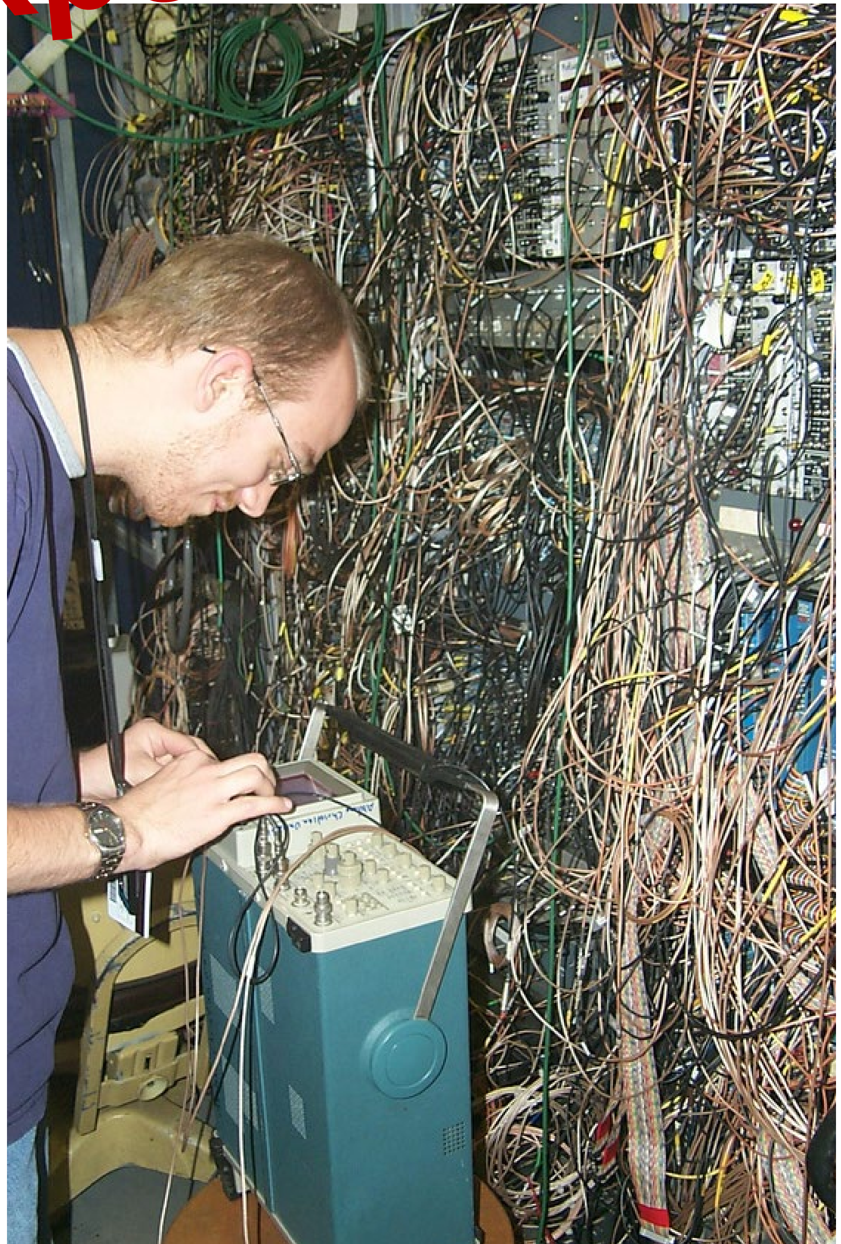
Research:

- Particle and Nuclear Physics
- Atom smashers
- Radiation Detectors
- Other interests: artificial intelligence, cosmology, amateur theology

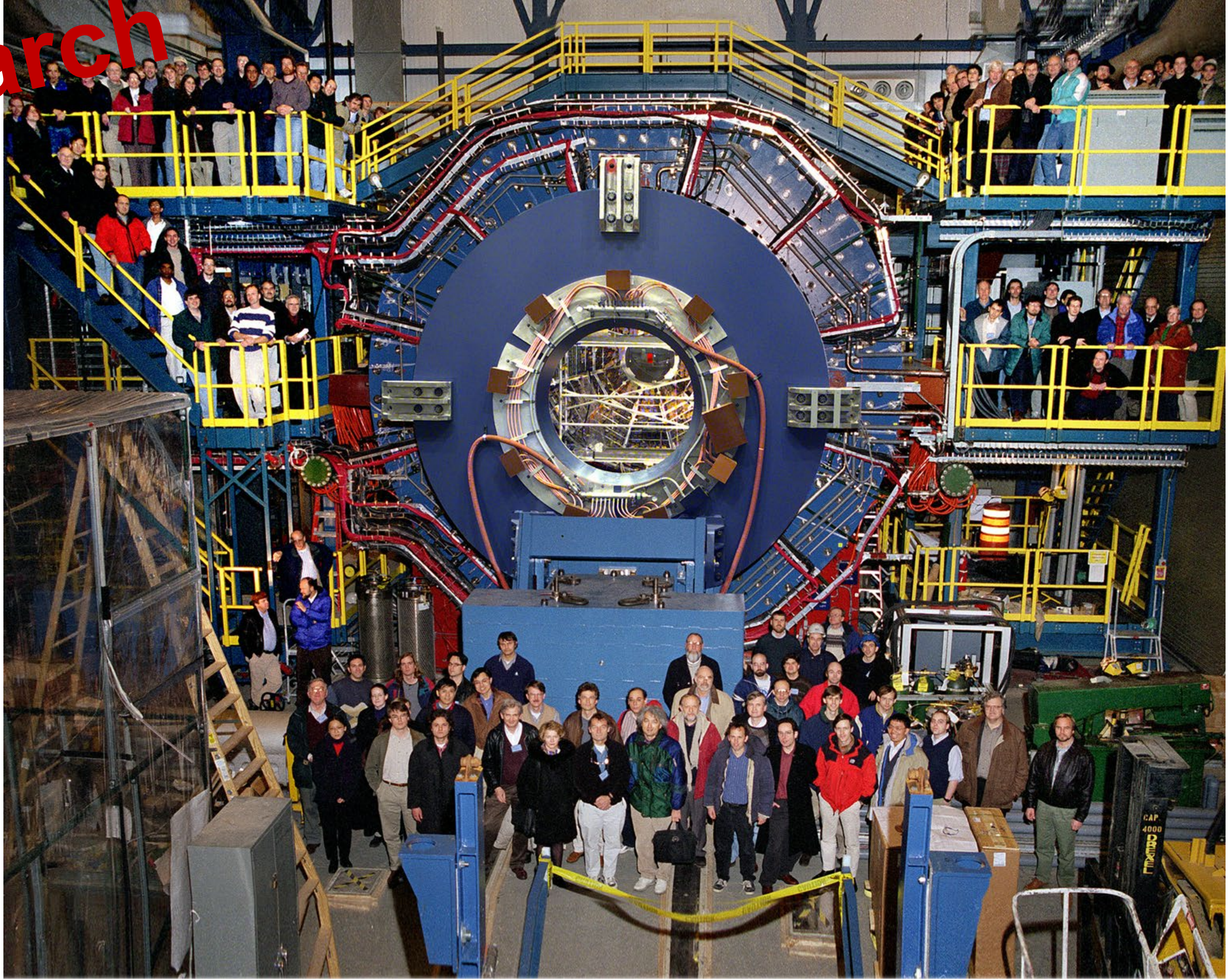
Hobbies:

- Eating
- Putting things in a laser
- Building dangerous things

First Experiment

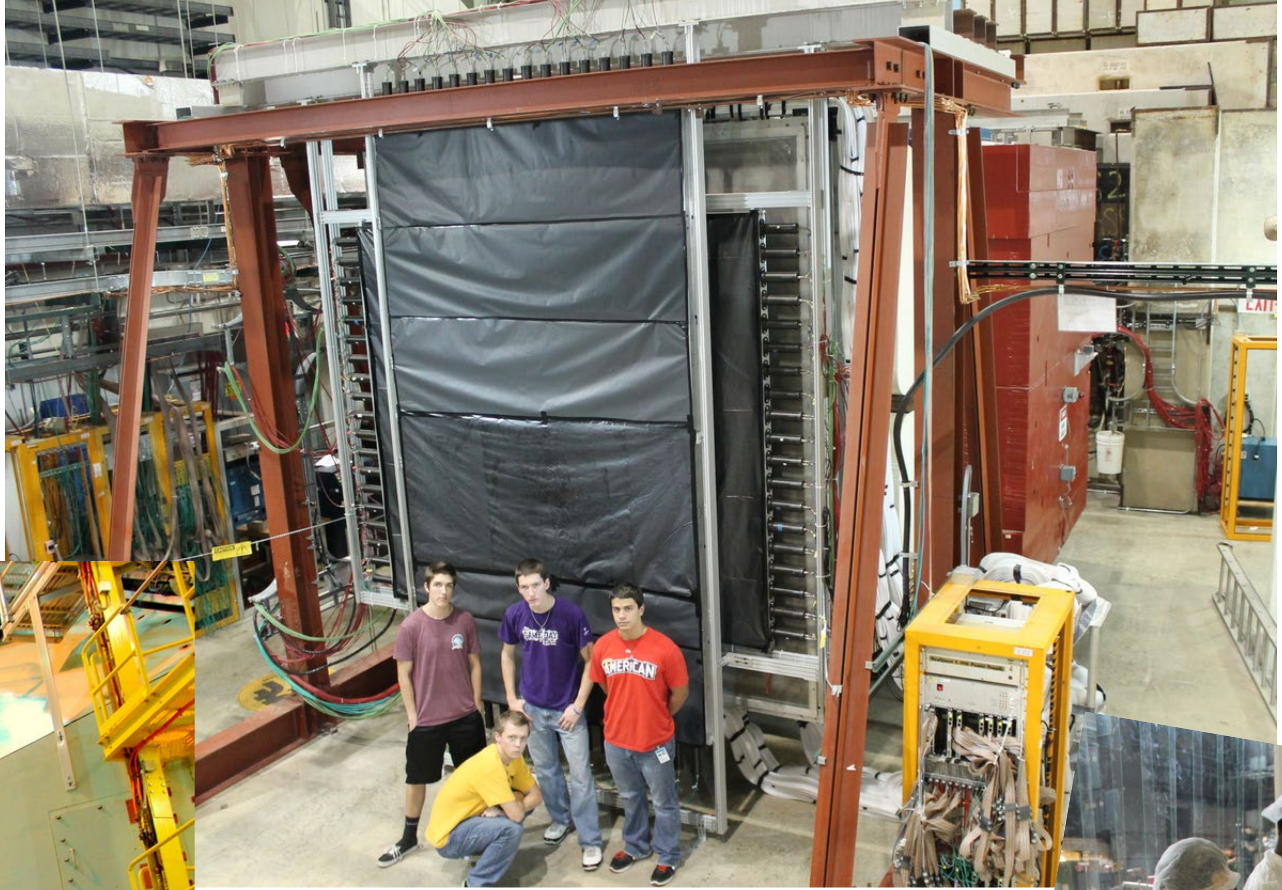


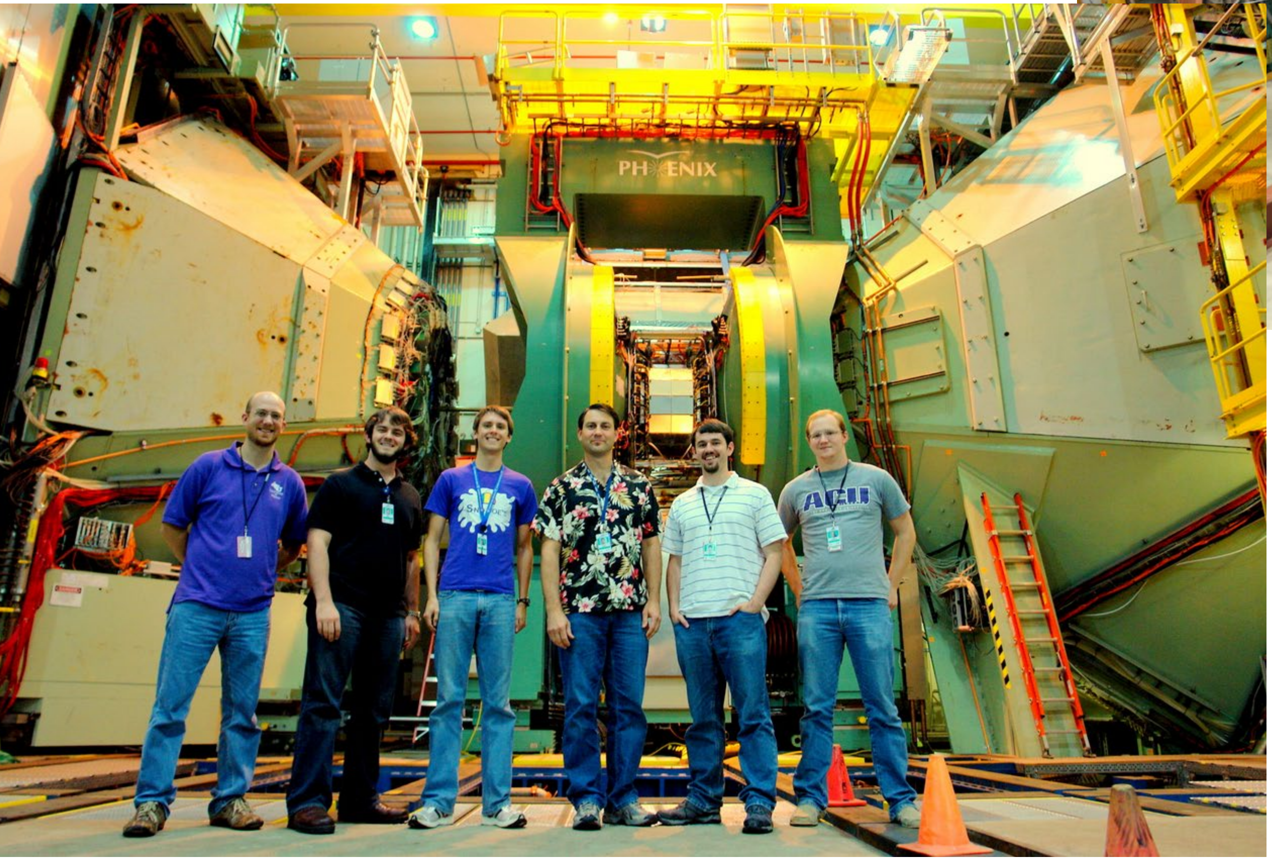
PhD Research



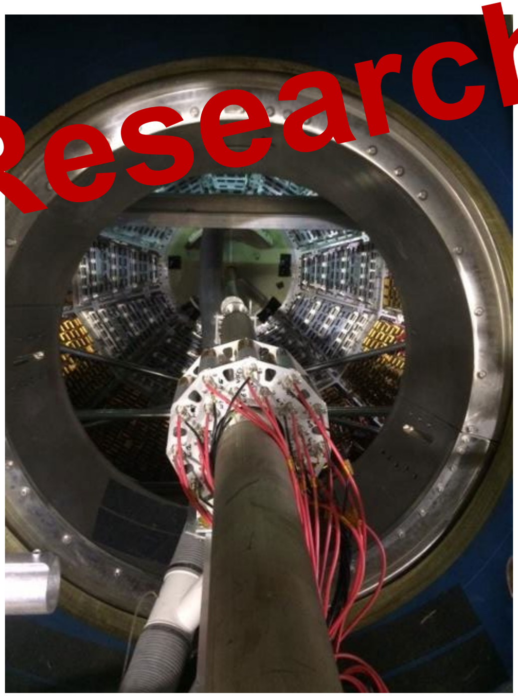
Mike Daugherity

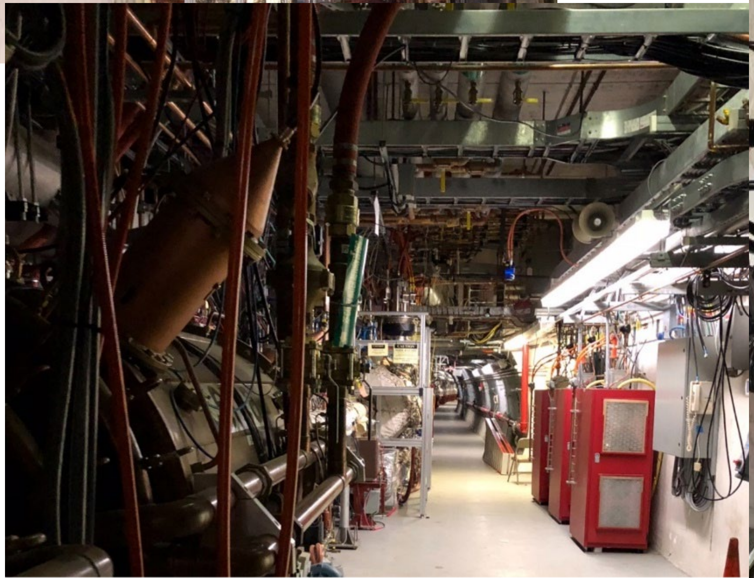
ACU Research

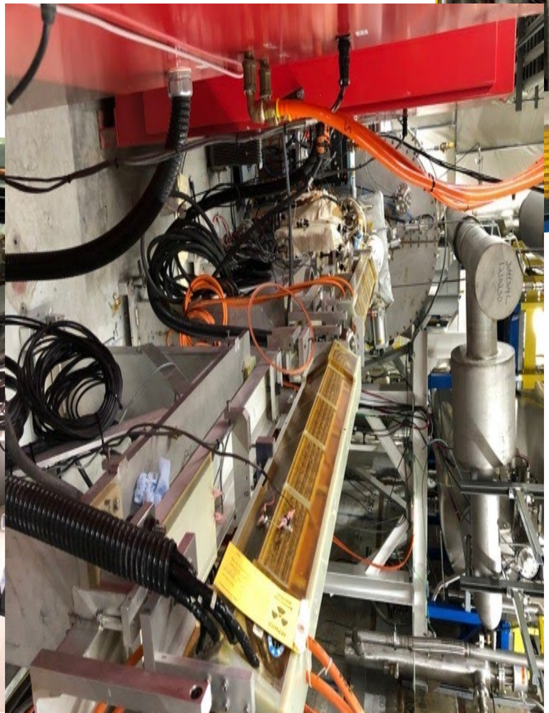




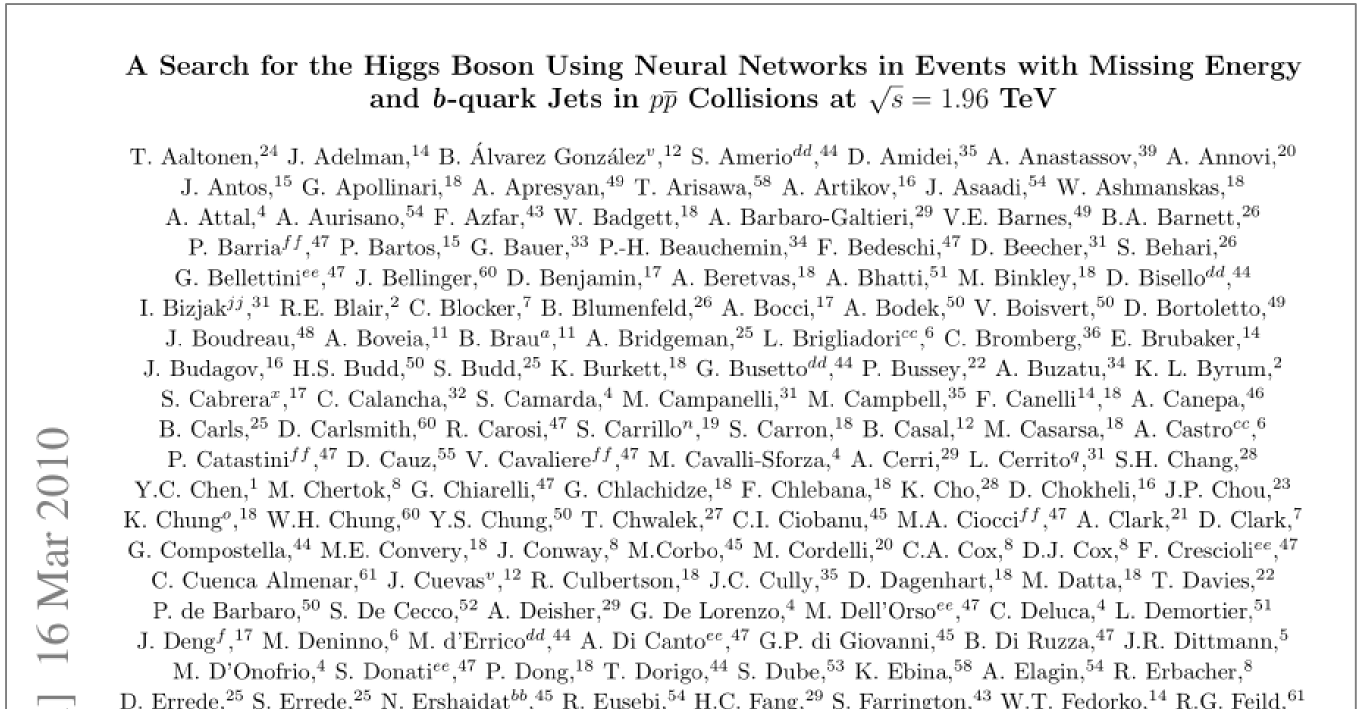
Summer Research







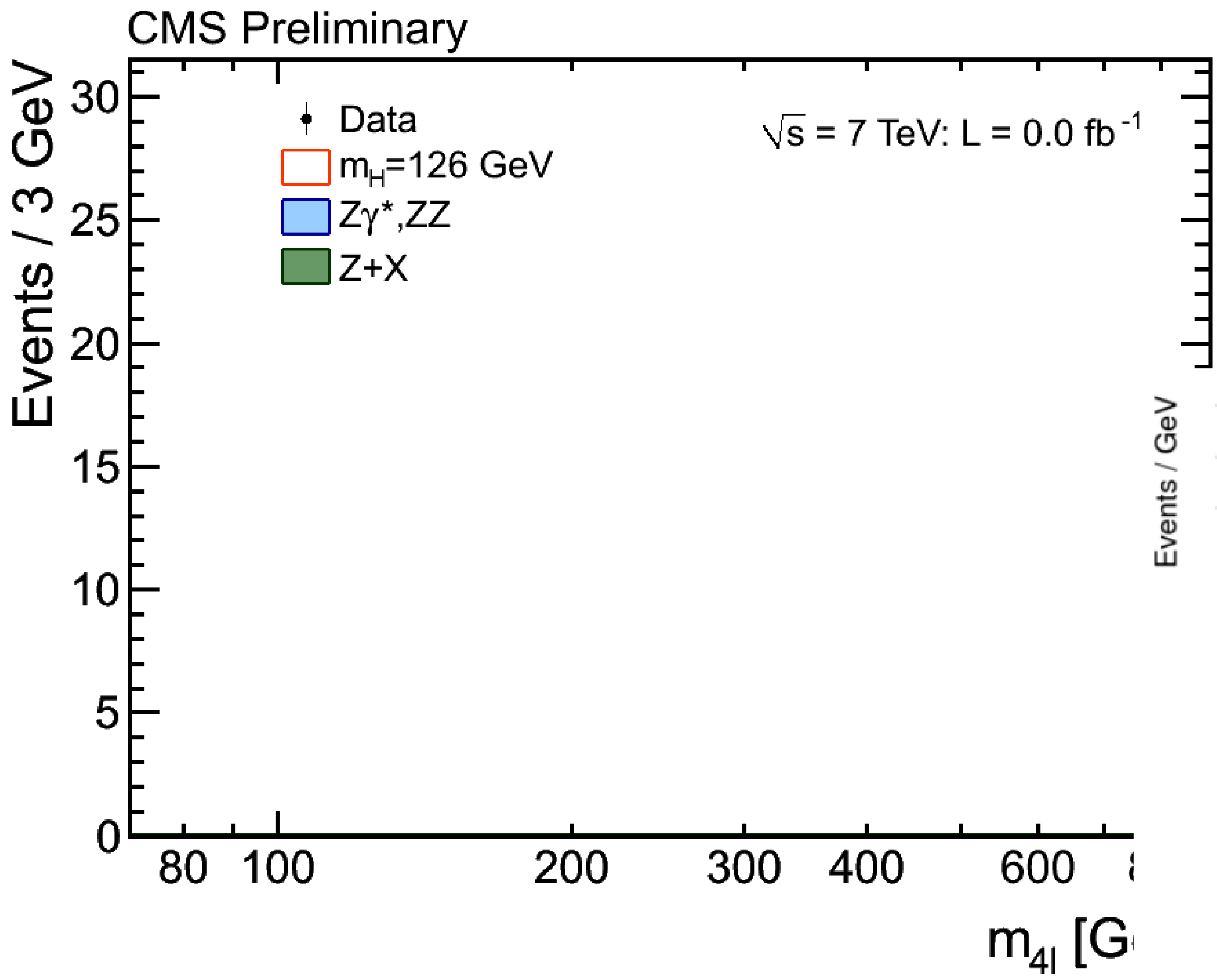
Before "Big Data" was a thing, there was high-energy physics…



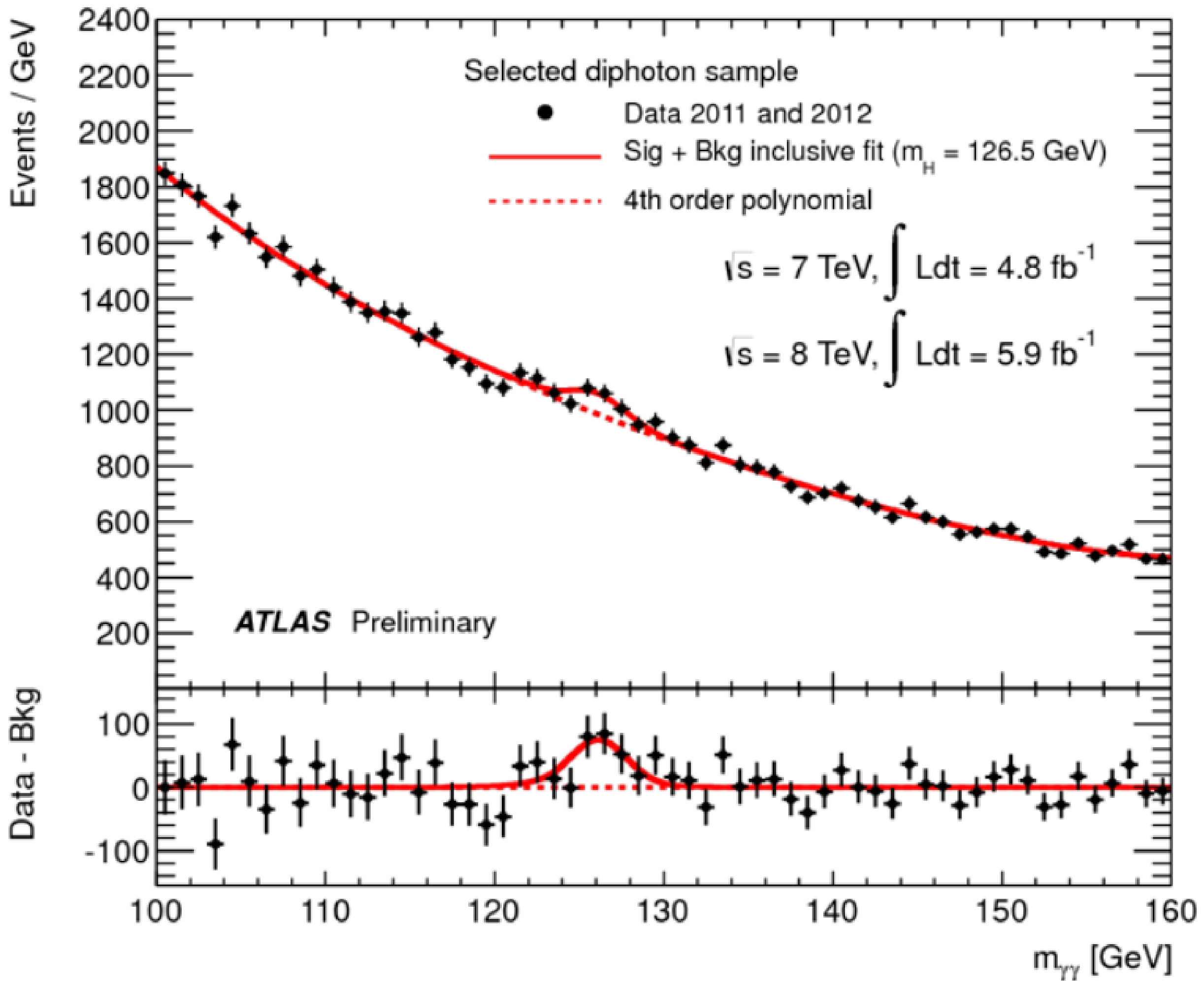
] 16 Mar 2010

A Search for the Higgs Boson Using Neural Networks in Events with Missing Energy and b-quark Jets in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV

T. Aaltonen,[24] J. Adelman,[14] B. Álvarez González[v,12] S. Amerio[dd,44] D. Amidei,[35] A. Anastassov,[39] A. Annovi,[20] J. Antos,[15] G. Apollinari,[18] A. Apresyan,[49] T. Arisawa,[58] A. Artikov,[16] J. Asaadi,[54] W. Ashmanskas,[18] A. Attal,[4] A. Aurisano,[54] F. Azfar,[43] W. Badgett,[18] A. Barbaro-Galtieri,[29] V.E. Barnes,[49] B.A. Barnett,[26] P. Barria[ff,47] P. Bartos,[15] G. Bauer,[33] P.-H. Beauchemin,[34] F. Bedeschi,[47] D. Beecher,[31] S. Behari,[26] G. Bellettini[ee,47] J. Bellinger,[60] D. Benjamin,[17] A. Beretvas,[18] A. Bhatti,[51] M. Binkley,[18] D. Bisello[dd,44] I. Bizjak[jj,31] R.E. Blair,[2] C. Blocker,[7] B. Blumenfeld,[26] A. Bocci,[17] A. Bodek,[50] V. Boisvert,[50] D. Bortoletto,[49] J. Boudreau,[48] A. Boveia,[11] B. Brau[a,11] A. Bridgeman,[25] L. Brigliadori[cc,6] C. Bromberg,[36] E. Brubaker,[14] J. Budagov,[16] H.S. Budd,[50] S. Budd,[25] K. Burkett,[18] G. Busetto[dd,44] P. Bussey,[22] A. Buzatu,[34] K. L. Byrum,[2] S. Cabrera[x,17] C. Calancha,[32] S. Camarda,[4] M. Campanelli,[31] M. Campbell,[35] F. Canelli[14,18] A. Canepa,[46] B. Carls,[25] D. Carlsmith,[60] R. Carosi,[47] S. Carrillo[n,19] S. Carron,[18] B. Casal,[12] M. Casarsa,[18] A. Castro[cc,6] P. Catastini[ff,47] D. Cauz,[55] V. Cavaliere[ff,47] M. Cavalli-Sforza,[4] A. Cerri,[29] L. Cerrito[q,31] S.H. Chang,[28] Y.C. Chen,[1] M. Chertok,[8] G. Chiarelli,[47] G. Chlachidze,[18] F. Chlebana,[18] K. Cho,[28] D. Chokheli,[16] J.P. Chou,[23] K. Chung[o,18] W.H. Chung,[60] Y.S. Chung,[50] T. Chwalek,[27] C.I. Ciobanu,[45] M.A. Ciocci[ff,47] A. Clark,[21] D. Clark,[7] G. Compostella,[44] M.E. Convery,[18] J. Conway,[8] M.Corbo,[45] M. Cordelli,[20] C.A. Cox,[8] D.J. Cox,[8] F. Crescioli[ee,47] C. Cuenca Almenar,[61] J. Cuevas[v,12] R. Culbertson,[18] J.C. Cully,[35] D. Dagenhart,[18] M. Datta,[18] T. Davies,[22] P. de Barbaro,[50] S. De Cecco,[52] A. Deisher,[29] G. De Lorenzo,[4] M. Dell'Orso[ee,47] C. Deluca,[4] L. Demortier,[51] J. Deng[f,17] M. Deninno,[6] M. d'Errico[dd,44] A. Di Canto[ee,47] G.P. di Giovanni,[45] B. Di Ruzza,[47] J.R. Dittmann,[5] M. D'Onofrio,[4] S. Donati[ee,47] P. Dong,[18] T. Dorigo,[44] S. Dube,[53] K. Ebina,[58] A. Elagin,[54] R. Erbacher,[8] D. Errede,[25] S. Errede,[25] N. Ershaidat[bb,45] R. Eusebi,[54] H.C. Fang,[29] S. Farrington,[43] W.T. Fedorko,[14] R.G. Feild,[61]



CMS Preliminary
Events / 3 GeV
Data
m_H=126 GeV
Zγ*,ZZ
Z+X
√s = 7 TeV: L = 0.0 fb⁻¹
m4l [G

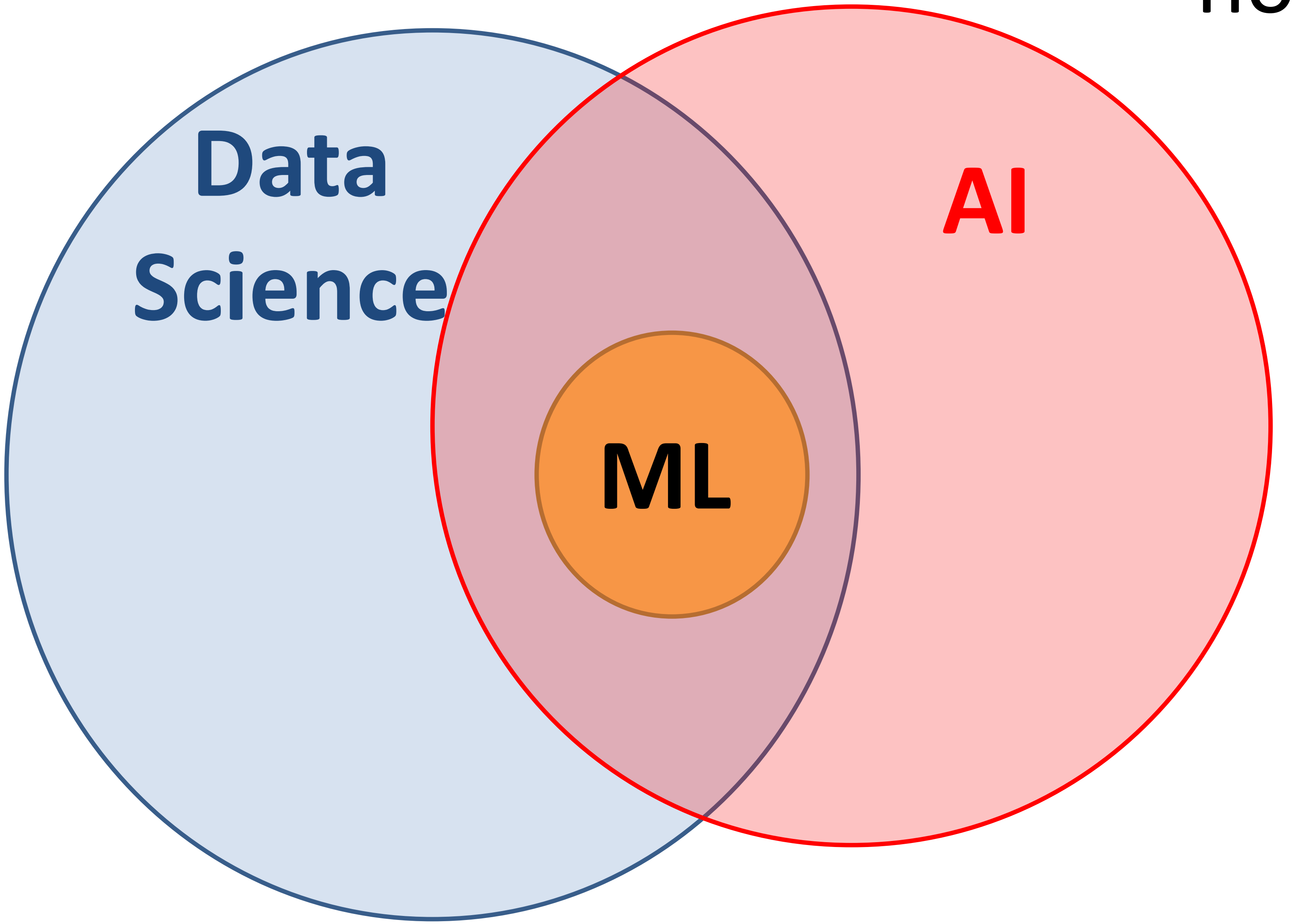


Events / GeV
Selected diphoton sample
Data 2011 and 2012
Sig + Bkg inclusive fit (m_H = 126.5 GeV)
4th order polynomial
√s = 7 TeV, ∫ Ldt = 4.8 fb⁻¹
√s = 8 TeV, ∫ Ldt = 5.9 fb⁻¹
ATLAS Preliminary
Data - Bkg
$m_{\gamma\gamma}$ [GeV]

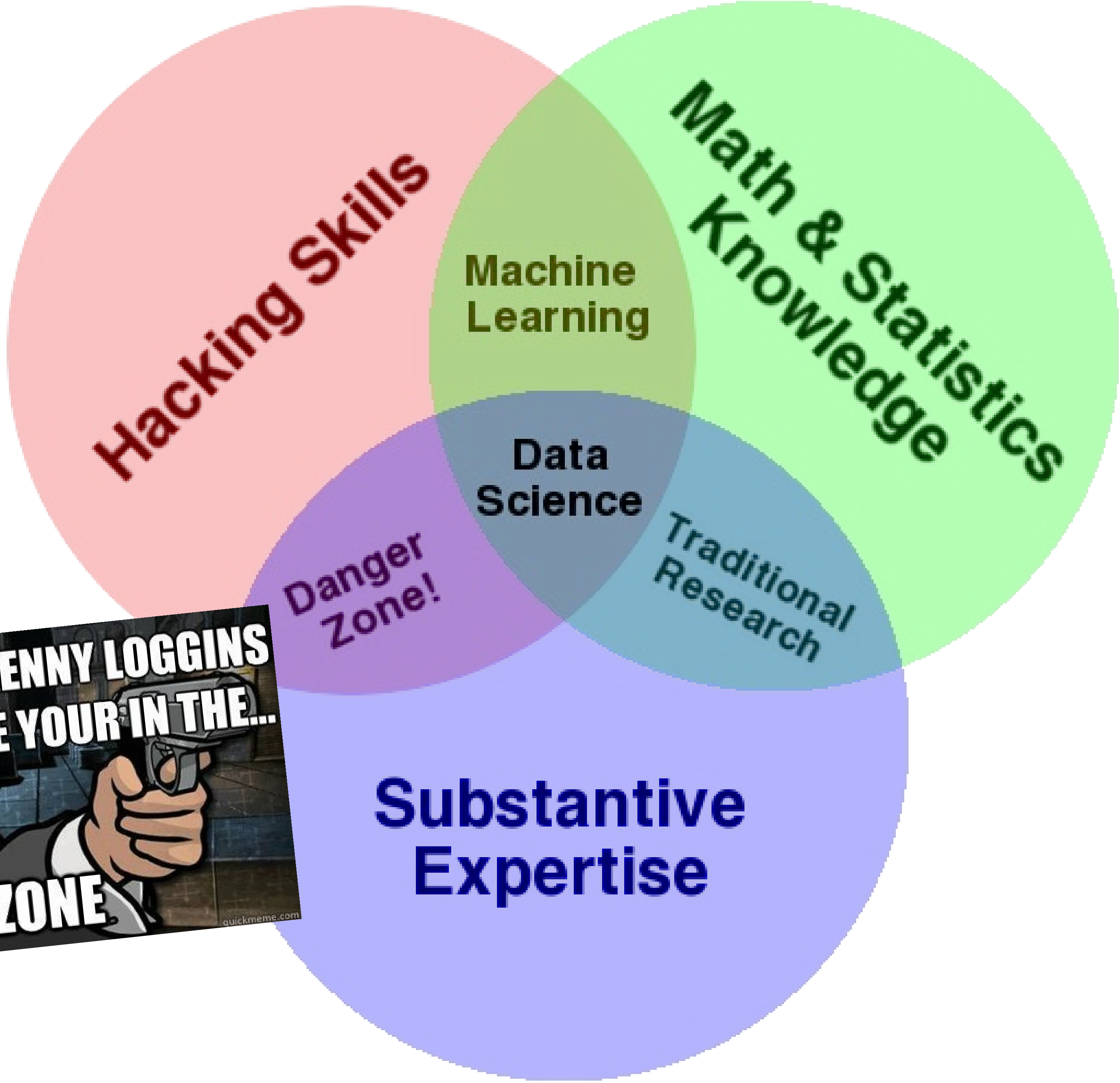
Part 2

WHAT IS PATTERN RECOGNITION: A TOO SHORT INTRO

* not to scale



Also: Pattern Recognition == Machine Learning


Hacking Skills
Math & Statistics Knowledge
Machine Learning
Data Science
Danger Zone!
Traditional Research
Substantive Expertise

CALL KENNY LOGGINS
CAUSE YOUR IN THE...
DANGER ZONE
quickmeme.com

http://mathbabe.org/2012/09/06/columbia-data-science-course-week-1-what-is-data-science/

"Data science, as it's practiced, is a blend of Red-Bull-fueled hacking and espresso-inspired statistics.

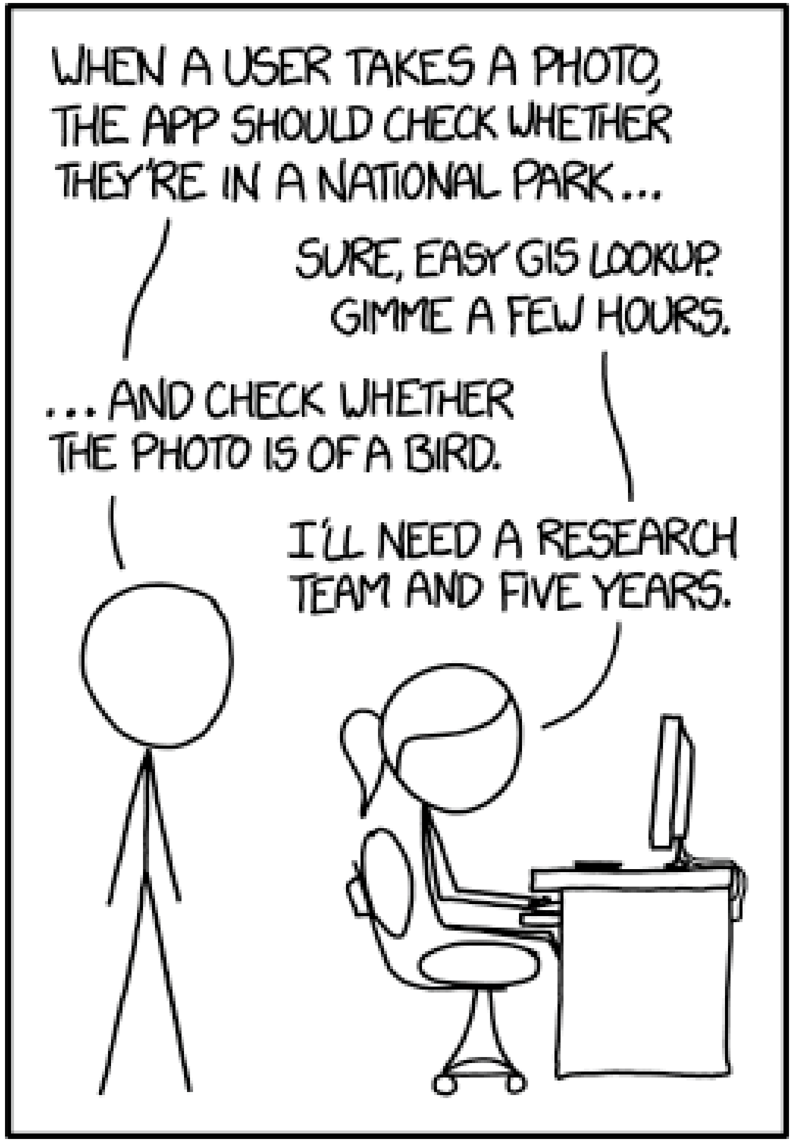
But data science is not merely hacking, because when hackers finish debugging their Bash one-liners and Pig scripts, few care about non-Euclidean distance metrics.

And data science is not merely statistics, because when statisticians finish theorizing the perfect model, few could read a tab delimited file into R if their job depended on it.

Data science is the civil engineering of data. Its acolytes possess a practical knowledge of tools & materials, coupled with a theoretical understanding of what's possible."

On One Hand

Relevant XKCD



IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

On The Other

Pattern Recognition

CLASSIFICATION

Spam Filter

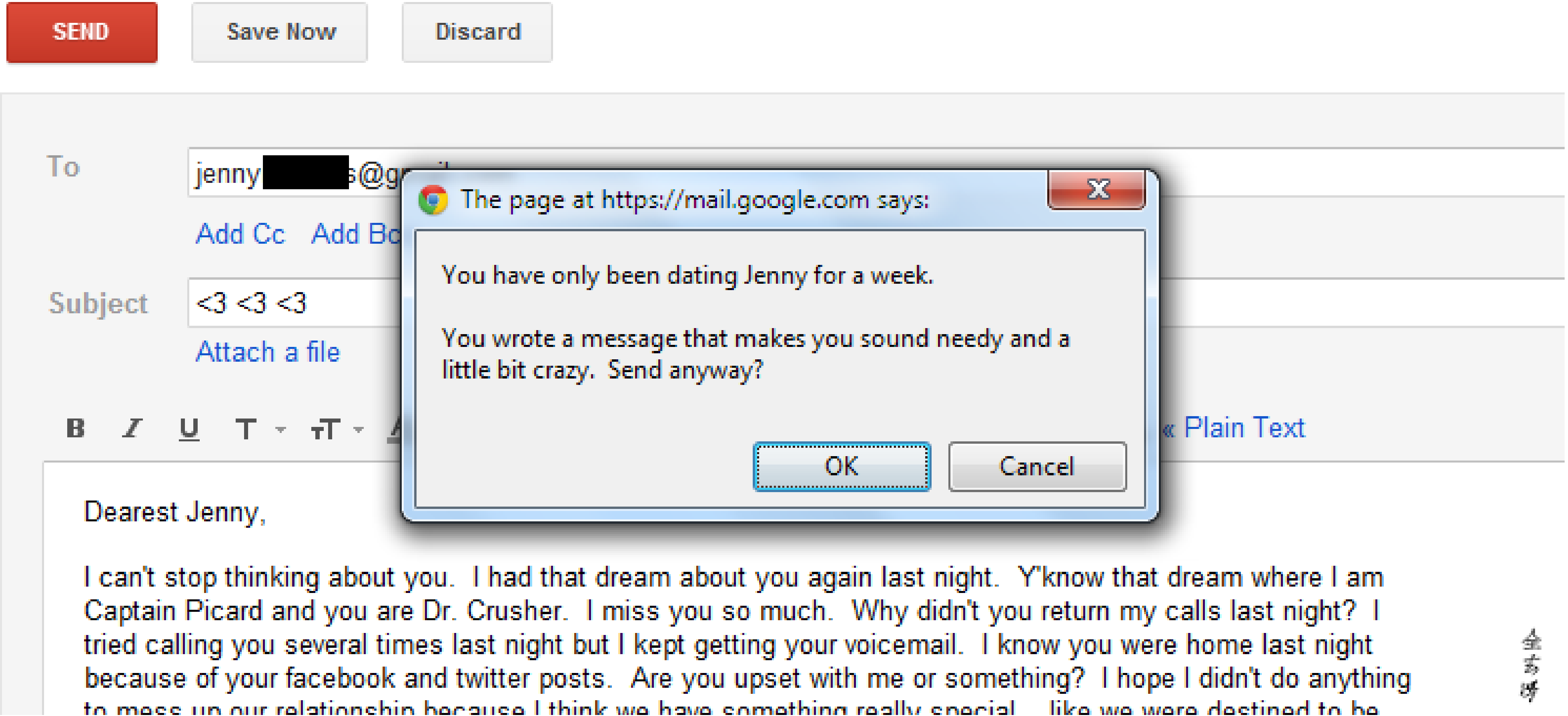
Assassinating spam e-mail

SpamAssassin is a widely used open-source spam filter. It calculates a score for an incoming e-mail, based on a number of built-in rules or ‘tests’ in SpamAssassin’s terminology, and adds a ‘junk’ flag and a summary report to the e-mail’s headers if the score is 5 or more.

```
-0.1 RCVD_IN_MXRATE_WL       RBL: MXRate recommends allowing
                             [123.45.6.789 listed in sub.mxrate.net]
 0.6 HTML_IMAGE_RATIO_02     BODY: HTML has a low ratio of text to image area
 1.2 TVD_FW_GRAPHIC_NAME_MID BODY: TVD_FW_GRAPHIC_NAME_MID
 0.0 HTML_MESSAGE            BODY: HTML included in message
 0.6 HTML_FONx_FACE_BAD      BODY: HTML font face is not a word
 1.4 SARE_GIF_ATTACH         FULL: Email has a inline gif
 0.1 BOUNCE_MESSAGE          MTA bounce message
 0.1 ANY_BOUNCE_MESSAGE      Message is some kind of bounce message
 1.4 AWL                     AWL: From: address is in the auto white-list
```

From left to right you see the score attached to a particular test, the test identifier, and a short description including a reference to the relevant part of the e-mail. As you see, scores for individual tests can be negative (indicating evidence suggesting the e-mail is ham rather than spam) as well as positive. The overall score of 5.3 suggests the e-mail might be spam.

Spam Filter


SEND
Save Now
Discard
To
jenny[redacted]s@g
Add Cc
Add Bc
Subject
<3 <3 <3
Attach a file
« Plain Text
The page at https://mail.google.com says:
You have only been dating Jenny for a week.
You wrote a message that makes you sound needy and a little bit crazy. Send anyway?
OK
Cancel
Dearest Jenny,
I can't stop thinking about you. I had that dream about you again last night. Y'know that dream where I am Captain Picard and you are Dr. Crusher. I miss you so much. Why didn't you return my calls last night? I tried calling you several times last night but I kept getting your voicemail. I know you were home last night because of your facebook and twitter posts. Are you upset with me or something? I hope I didn't do anything

Kinect

Real-Time Human Pose Recognition in Parts from Single Depth Images

Jamie Shotton Andrew Fitzgibbon Mat Cook Toby Sharp Mark Finocchio
Richard Moore Alex Kipman Andrew Blake
Microsoft Research Cambridge & Xbox Incubation

Abstract

We propose a new method to quickly and accurately predict 3D positions of body joints from a single depth image, using no temporal information. We take an object recognition approach, designing an intermediate body parts representation that maps the difficult pose estimation problem into a simpler per-pixel classification problem. Our large and highly varied training dataset allows the classifier to estimate body parts invariant to pose, body shape, clothing, etc. Finally we generate confidence-scored 3D proposals of several body joints by reprojecting the classification result and finding local modes.

The system runs at 200 frames per second on consumer hardware. Our evaluation shows high accuracy on both synthetic and real test sets, and investigates the effect of several training parameters. We achieve state of the art accuracy in our comparison with related work and demonstrate improved generalization over exact whole-skeleton nearest neighbor matching.

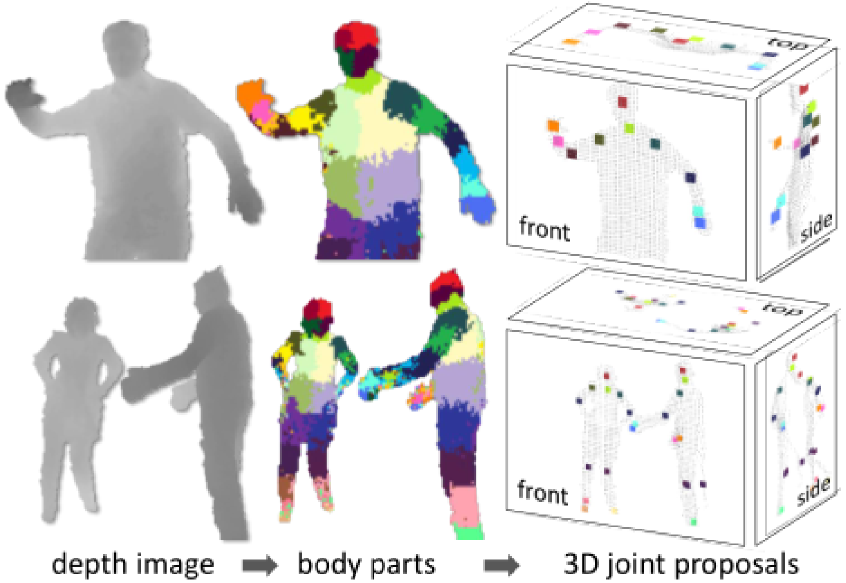
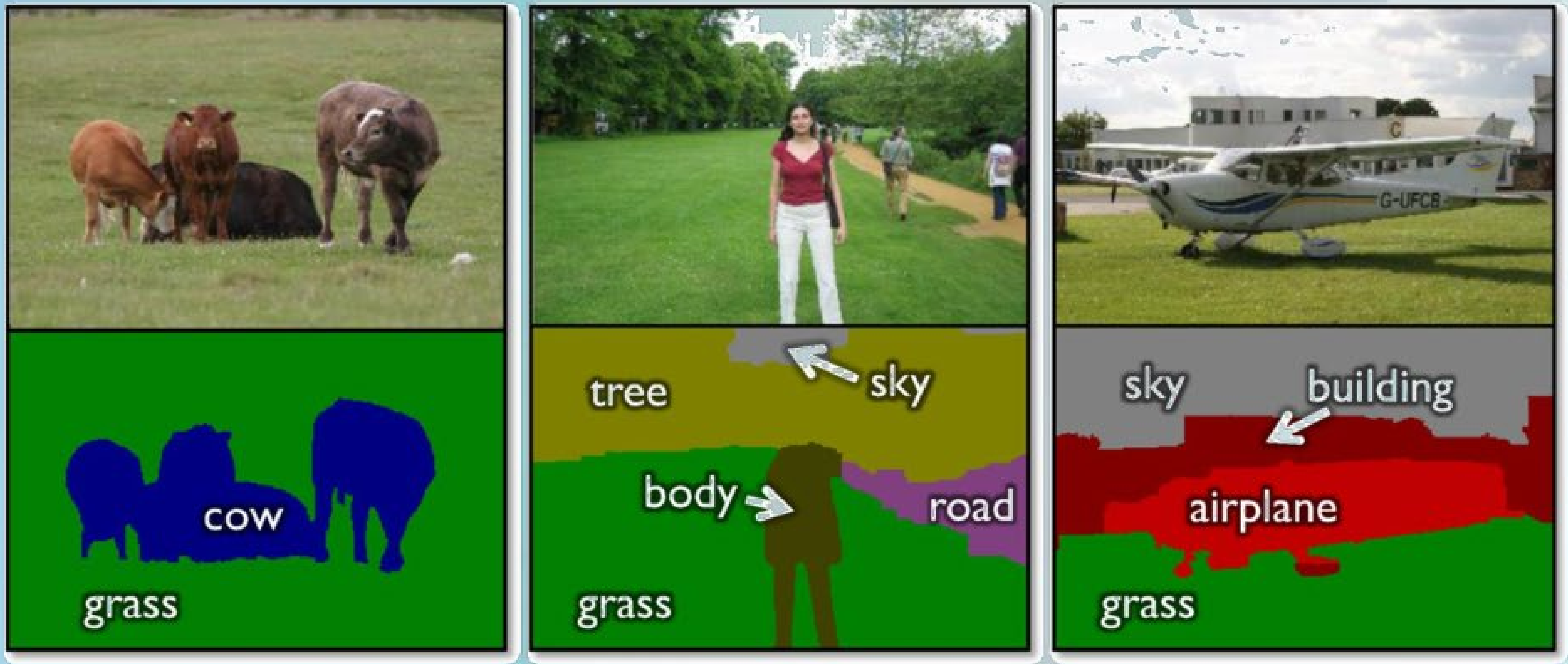


Figure 1. **Overview.** From an single input depth image, a per-pixel body part distribution is inferred. (Colors indicate the most likely part labels at each pixel, and correspond in the joint proposals). Local modes of this signal are estimated to give high-quality proposals for the 3D locations of body joints, even for multiple users.

Aside: Object Recognition



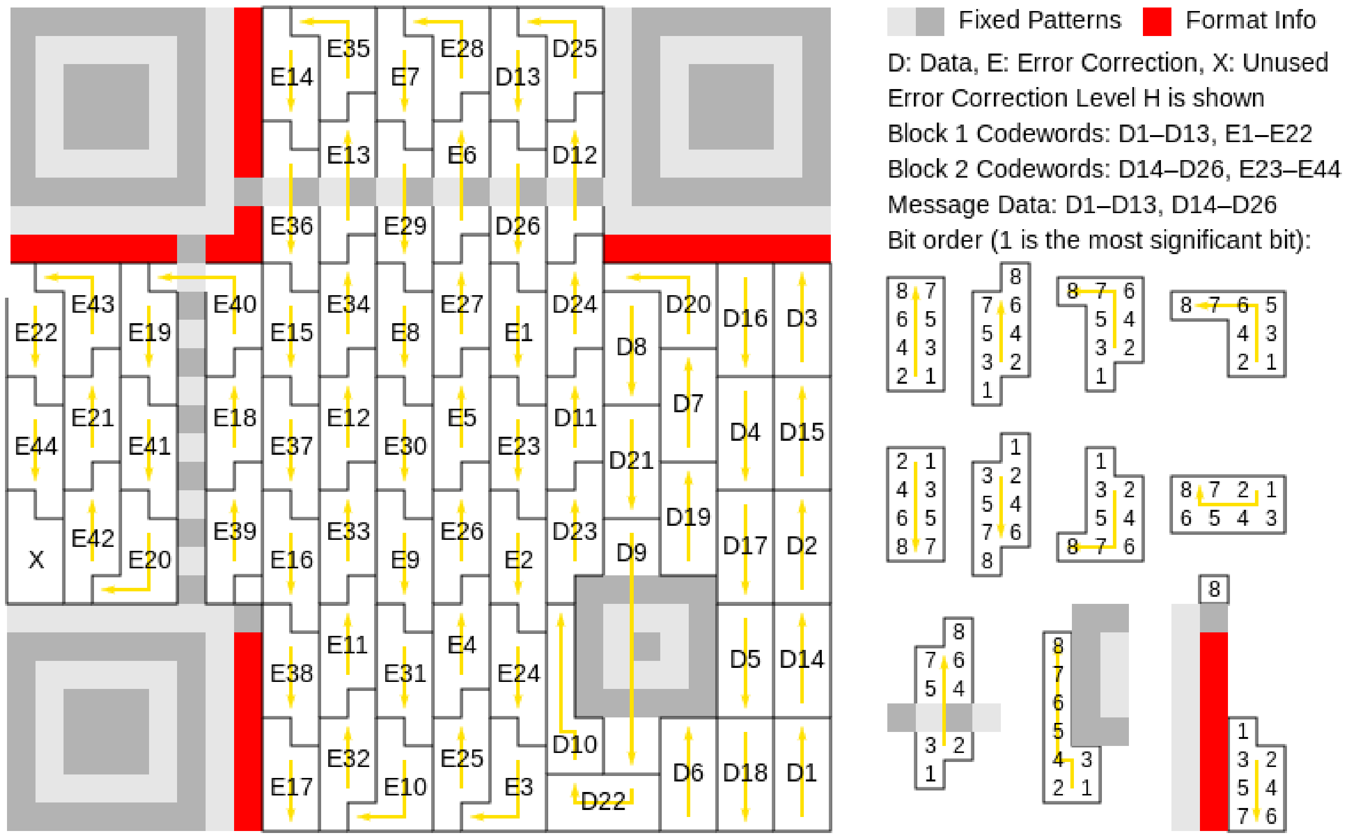
object classes	building	grass	tree	cow	sheep	sky	airplane	water	face	car
bicycle	flower	sign	bird	book	chair	road	cat	dog	body	boat

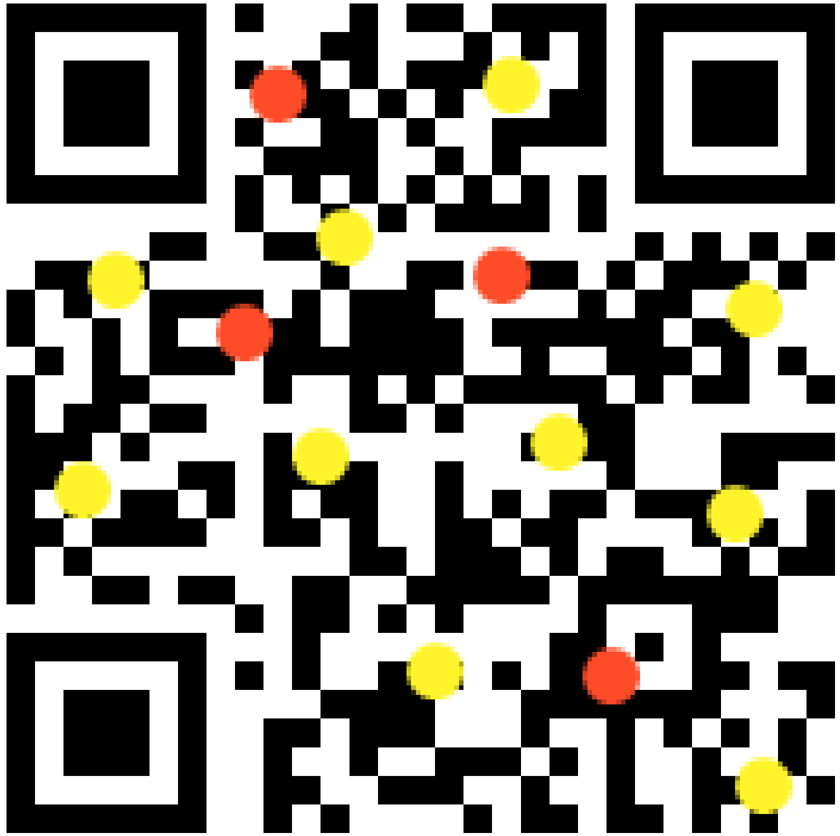
J. Shotton, J. Winn, C. Rother, A. Criminisi, ***TextonBoost*: Joint Appearance, Shape and Context Modeling for Multi-Class Object Recognition and Segmentation.** European Conference on Computer Vision, 2006

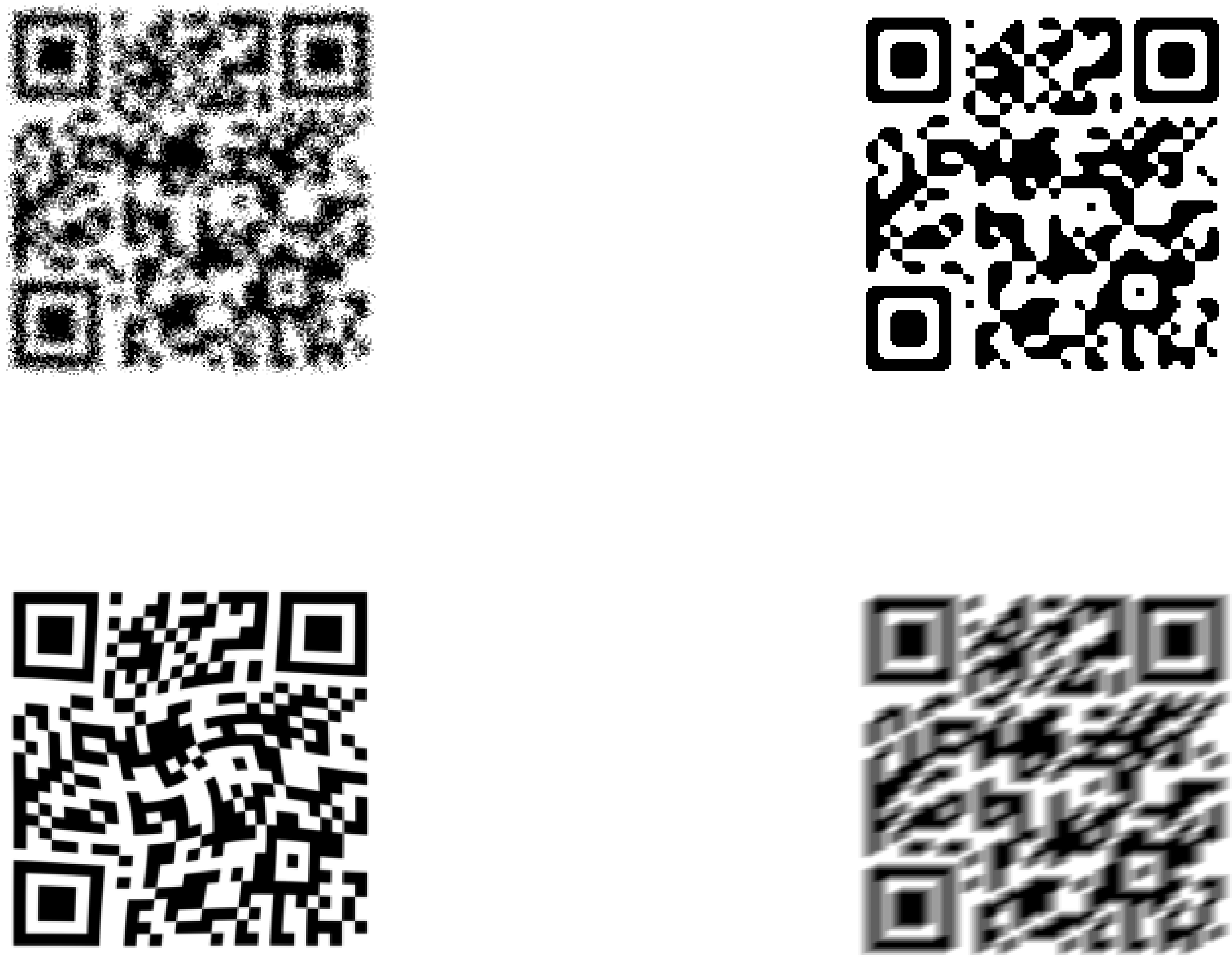
EXAMPLE

WOUNDED QR CODES

http://www.datagenetics.com/blog/november12013/index.html







Word Lens

https://www.youtube.com/watch?v=h2OfQdYrHRs
https://www.youtube.com/watch?v=zSOlYdlqyTQ

Other Links

- https://www.netsafe.org.nz/rescam/
- http://detexify.kirelabs.org/classify.html
- https://www.youtube.com/watch?v=qv6UVOQ0F44 – MARI/O
- https://www.youtube.com/watch?v=DMyJ6fGUqRI&t=46s – Rock Wall Pong
- https://xkcd.com/2236/ - Is it Christmas?

Pattern Recognition

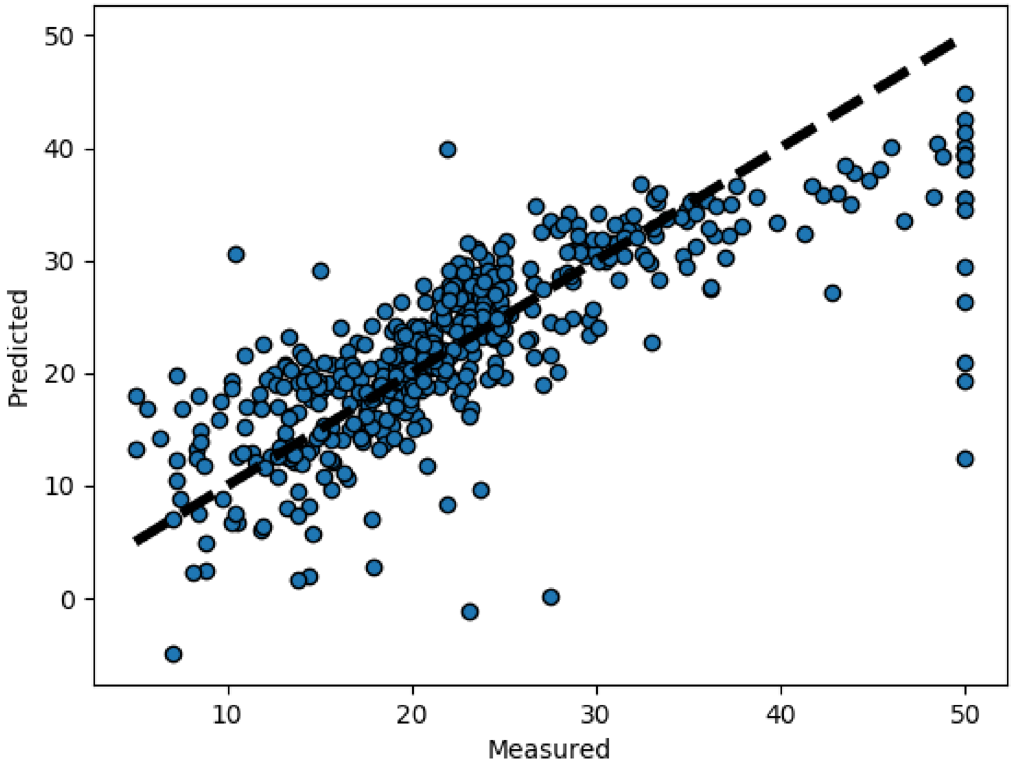
REGRESSION

Regression

Examples:

- predict the price of a house in Boston
- predict how much money a movie will make
- guess the value of a stock
- Diabetes dataset: predict measure of disease based on patient data

http://scikit-learn.org/stable/auto_examples/plot_cv_predict.html



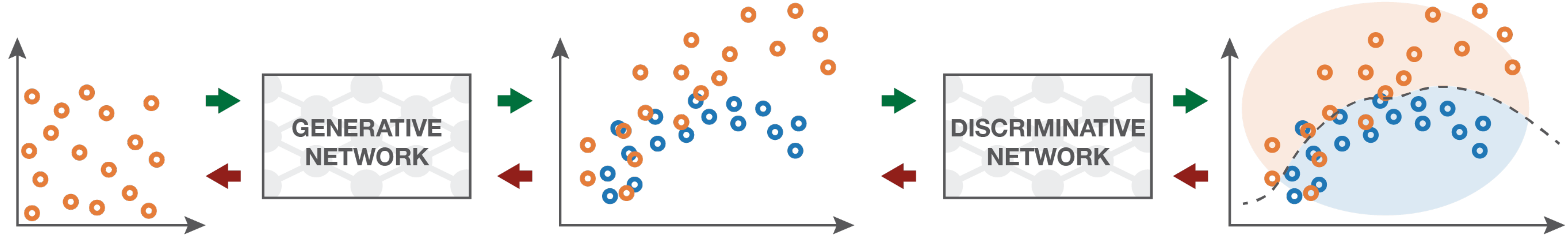

```
from sklearn import datasets
from sklearn.model_selection import cross_val_predict
from sklearn import linear_model
import matplotlib.pyplot as plt

lr = linear_model.LinearRegression()
boston = datasets.load_boston()
y = boston.target

# cross_val_predict returns an array of the same size as `y` where each entry
# is a prediction obtained by cross validation:
predicted = cross_val_predict(lr, boston.data, y, cv=10)

fig, ax = plt.subplots()
ax.scatter(y, predicted, edgecolors=(0, 0, 0))
ax.plot([y.min(), y.max()], [y.min(), y.max()], 'k--', lw=4)
ax.set_xlabel('Measured')
ax.set_ylabel('Predicted')
plt.show()
```

Pattern Recognition

GENERATION

Forward propagation (generation and classification)

Backward propagation (adversarial training)



Input random variables.

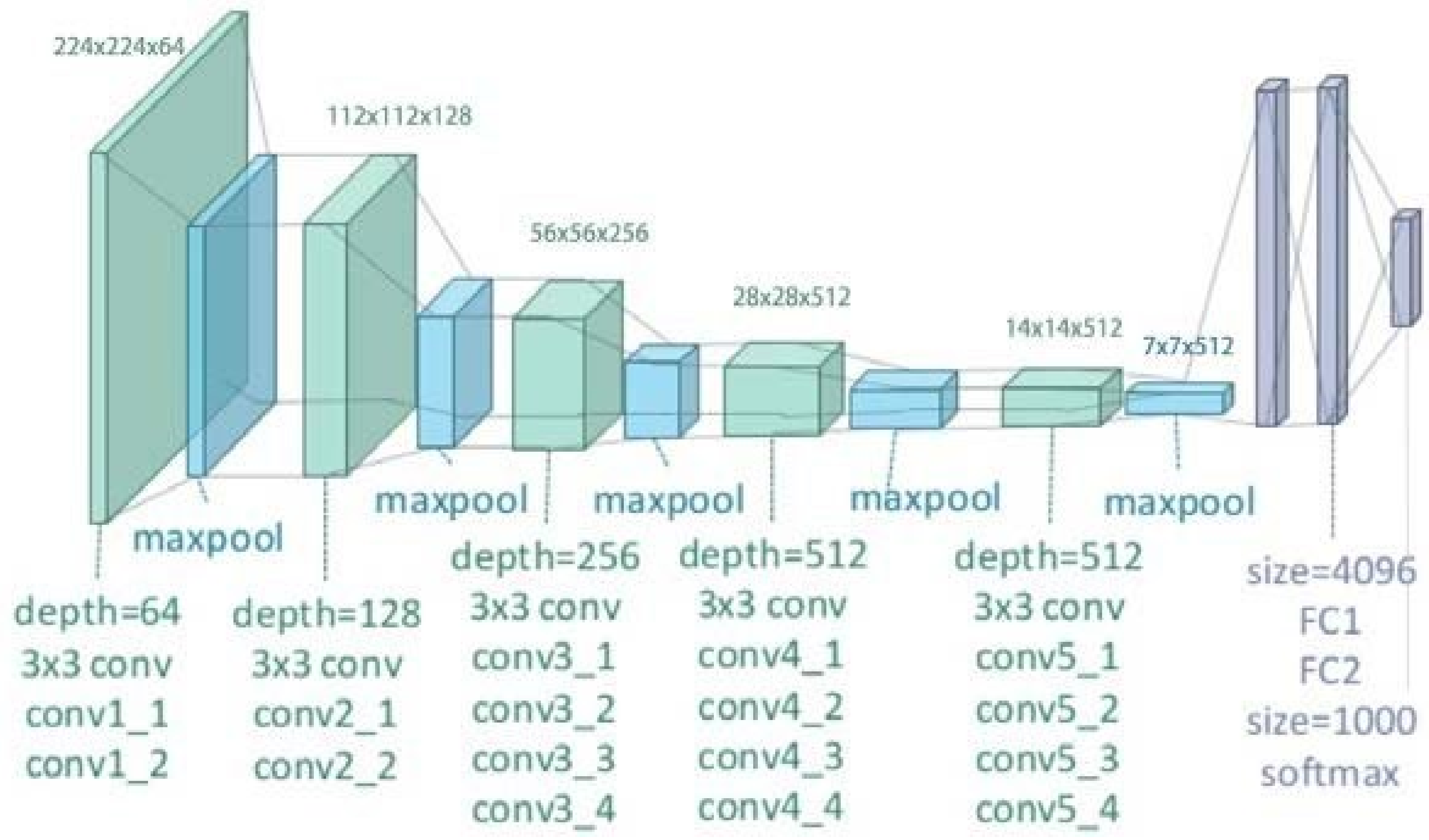
The generative network is trained to **maximise** the final classification error.

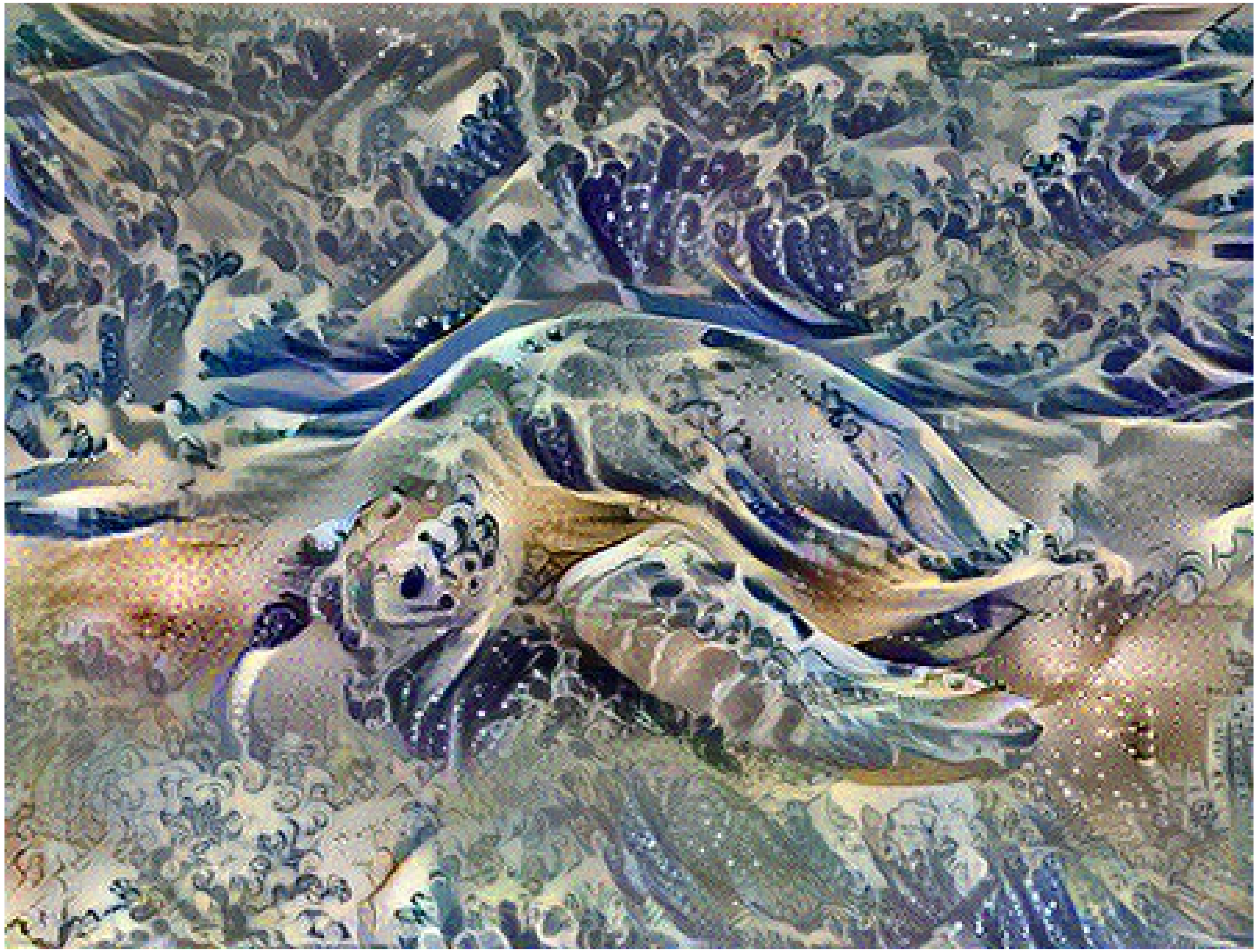
The generated distribution and the true distribution are not compared directly.

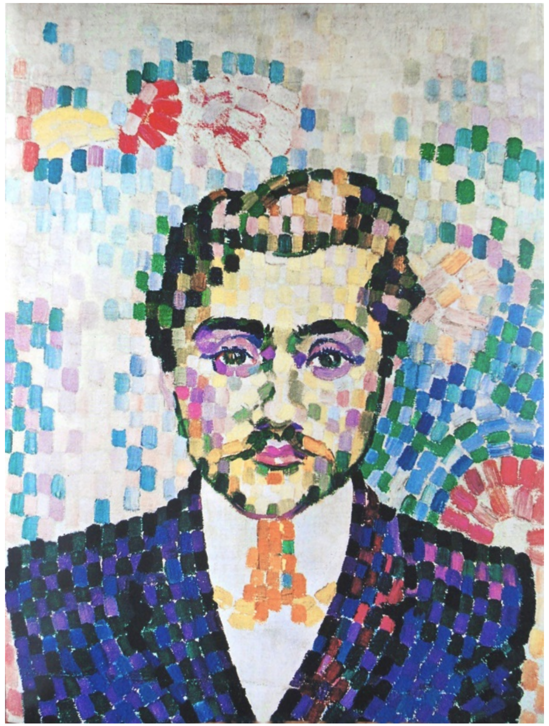
The discriminative network is trained to **minimise** the final classification error.

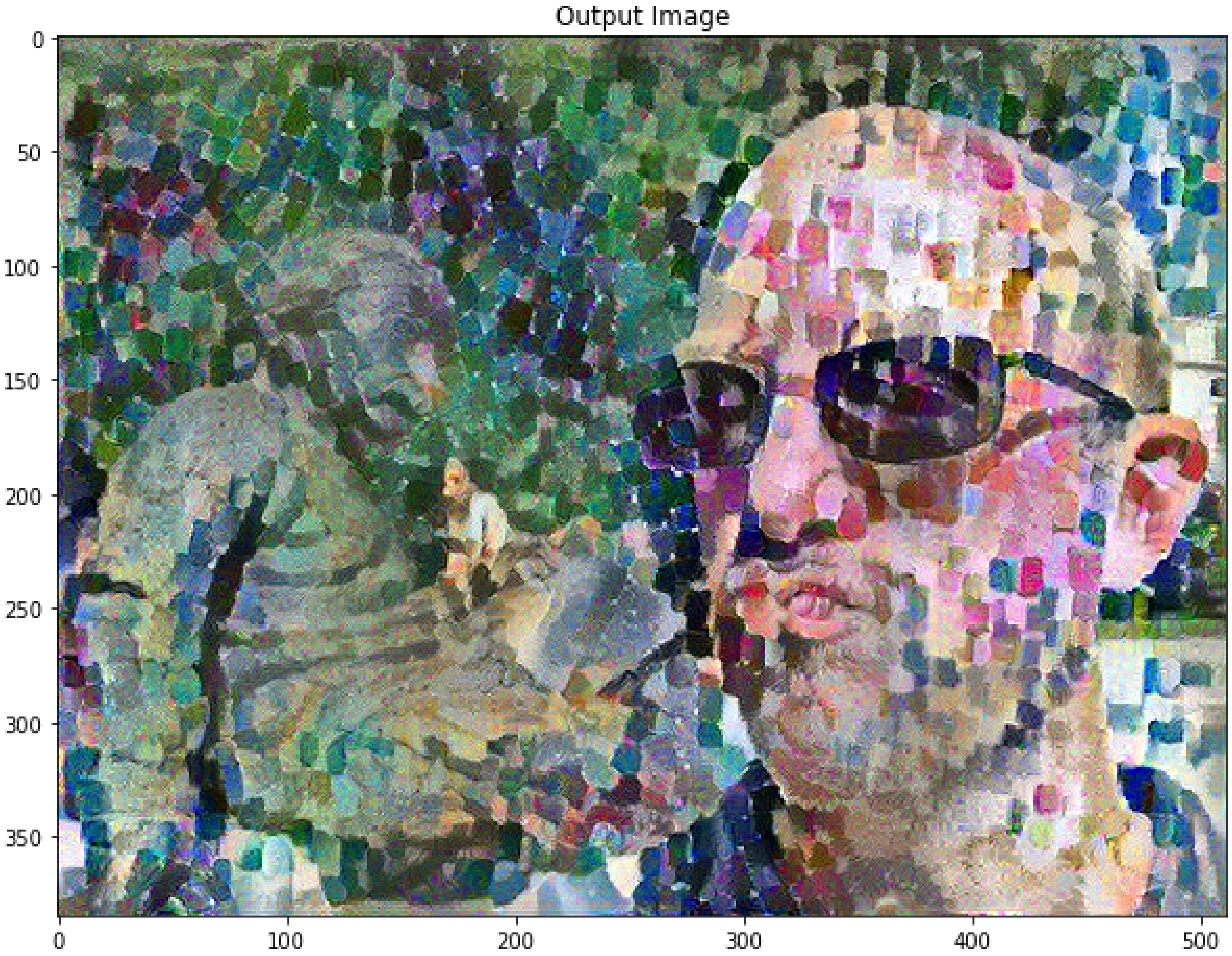
The classification error is the basis metric for the training of both networks.

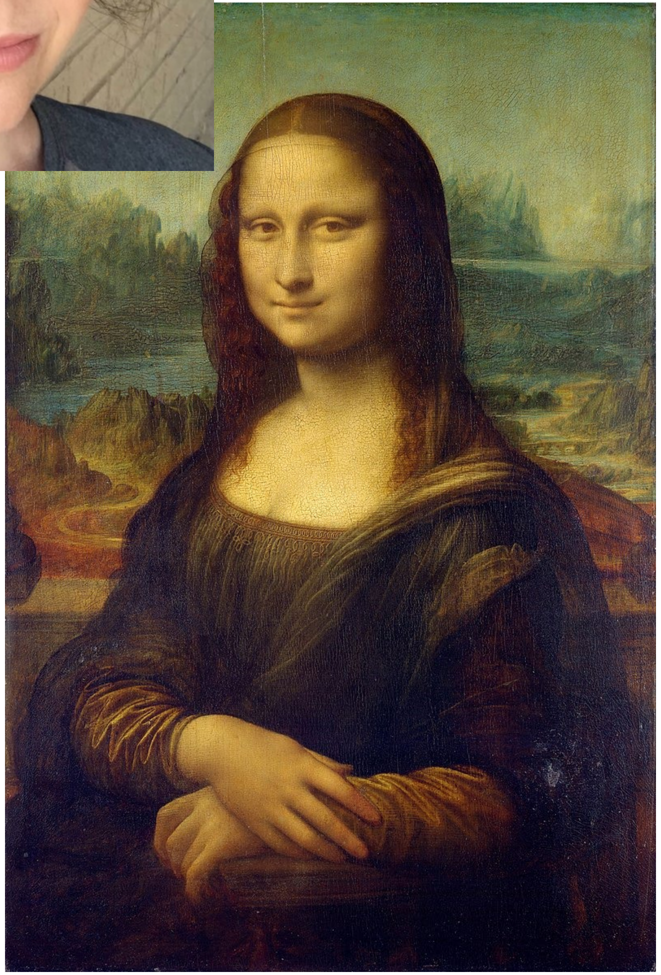
https://colab.research.google.com/github/tensorflow/models/blob/master/research/nst_blogpost/4_Neural_Style_Transfer_with_Eager_Execution.ipynb

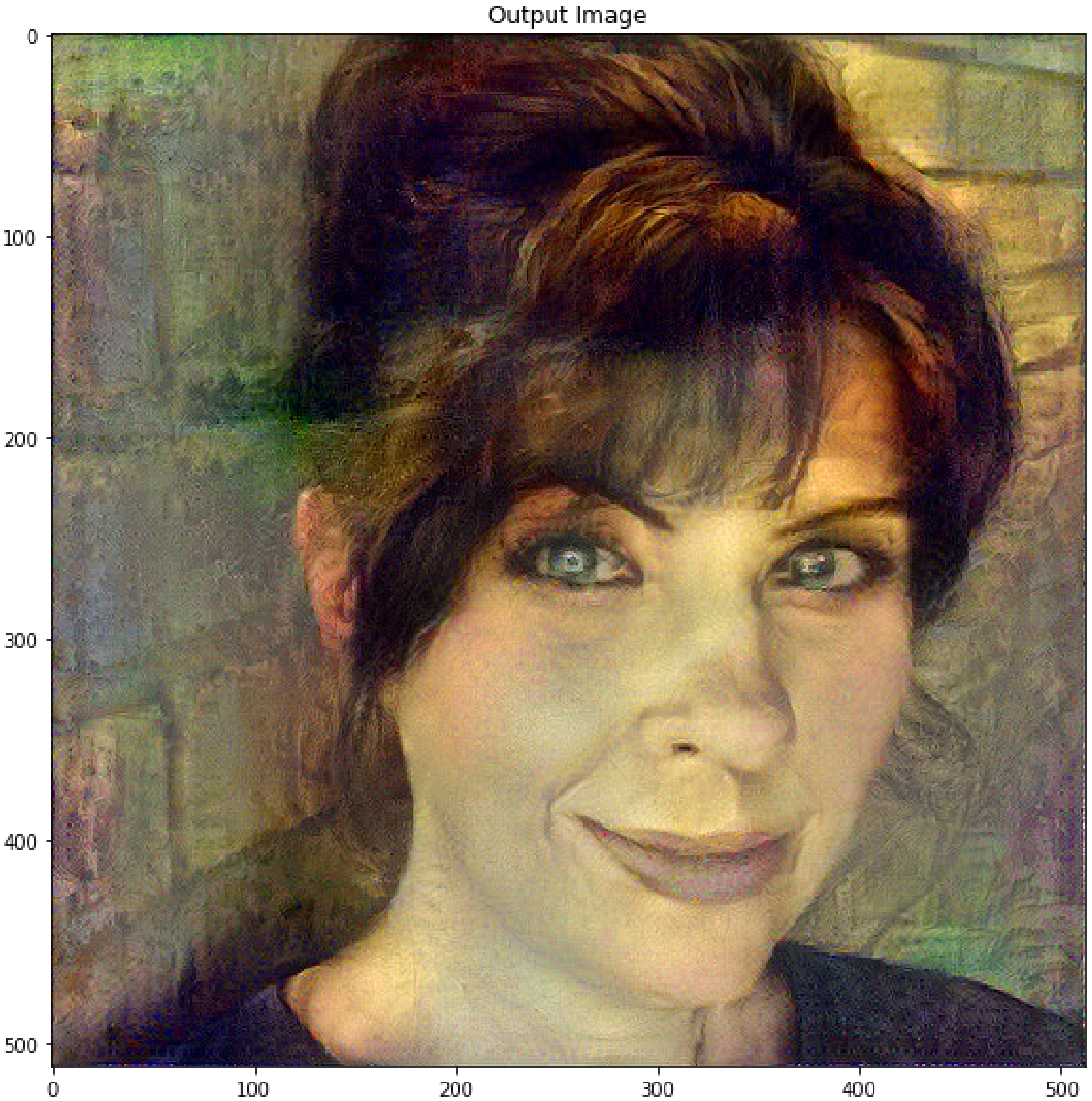






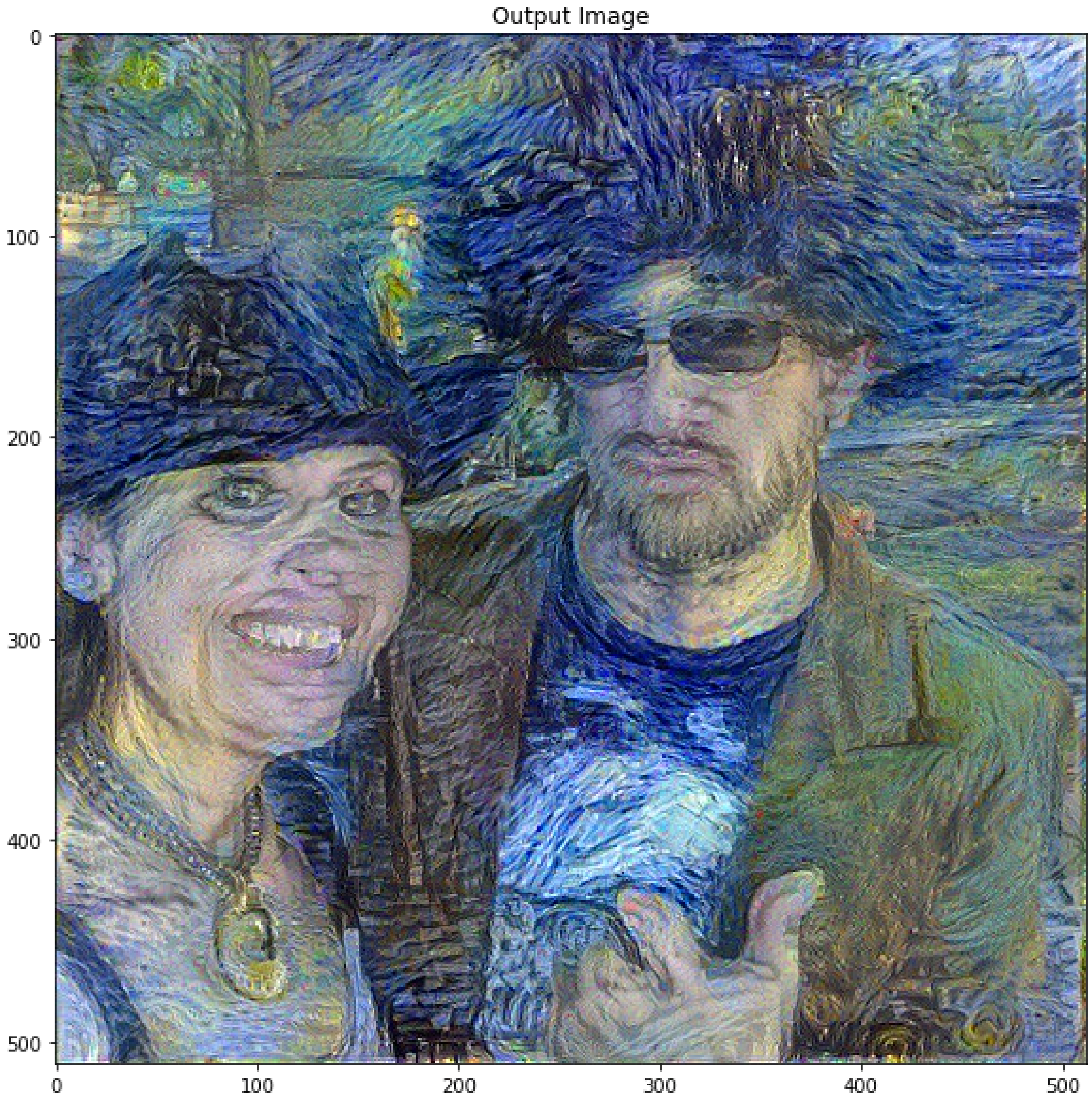

Output Image
0
50
100
150
200
250
300
350
0
100
200
300
400
500




Output Image
0
100
200
300
400
500
0
100
200
300
400
500


Output Image
0
100
200
300
400
500
0
100
200
300
400
500


Output Image
0
50
100
150
200
250
300
350
0
100
200
300
400
500


Output Image
0
100
200
300
400
500
0
100
200
300
400
500

What is Pattern Recognition?

Your turn:

My turn:

Make a decision based on data

i.e. find “patterns” in the data. Besides, if you aren’t making your decisions based on data, then what are you basing them on?

To get a good answer you MUST:

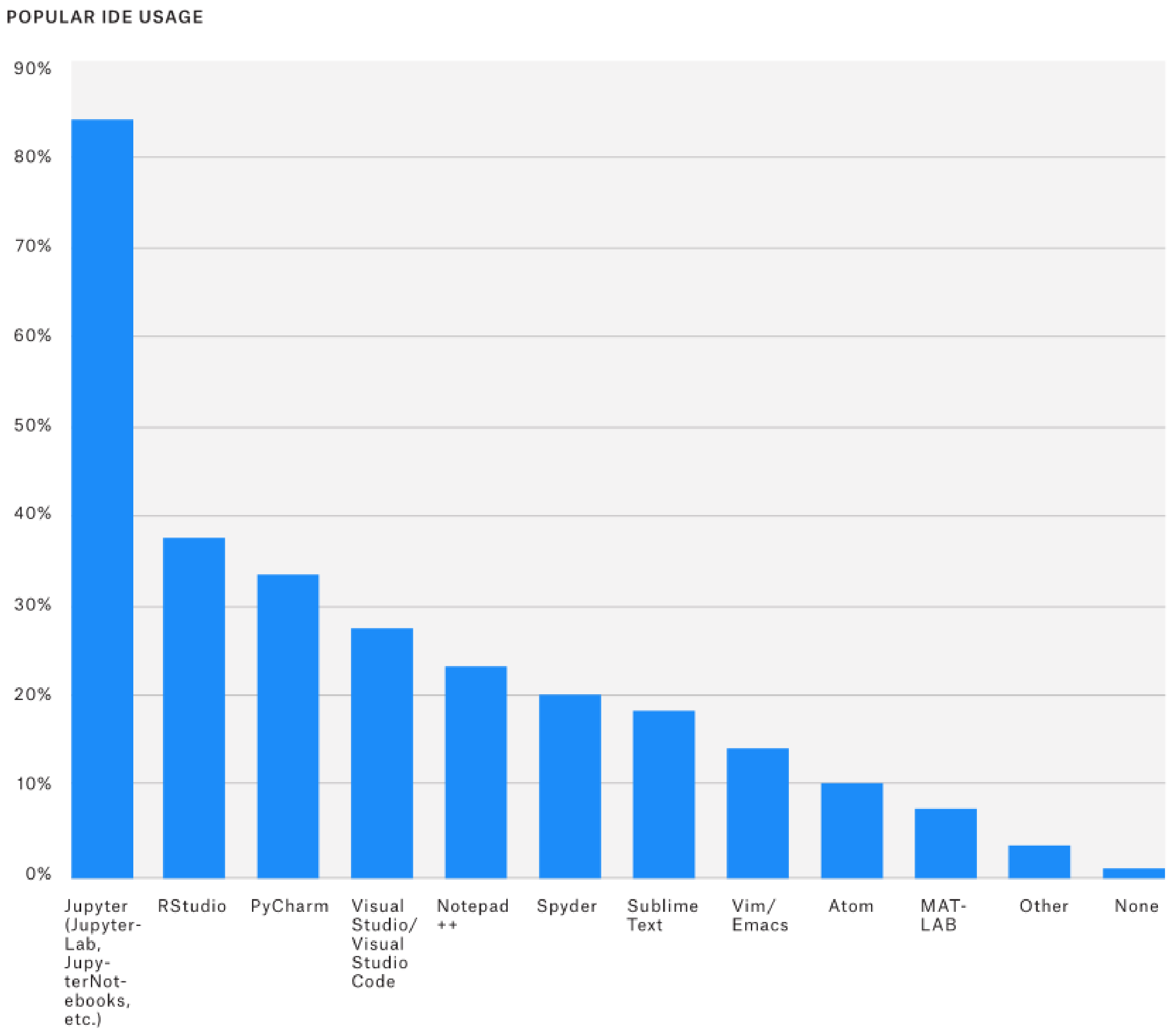
- ask a good question
- have good data

Part 3

TOOLS

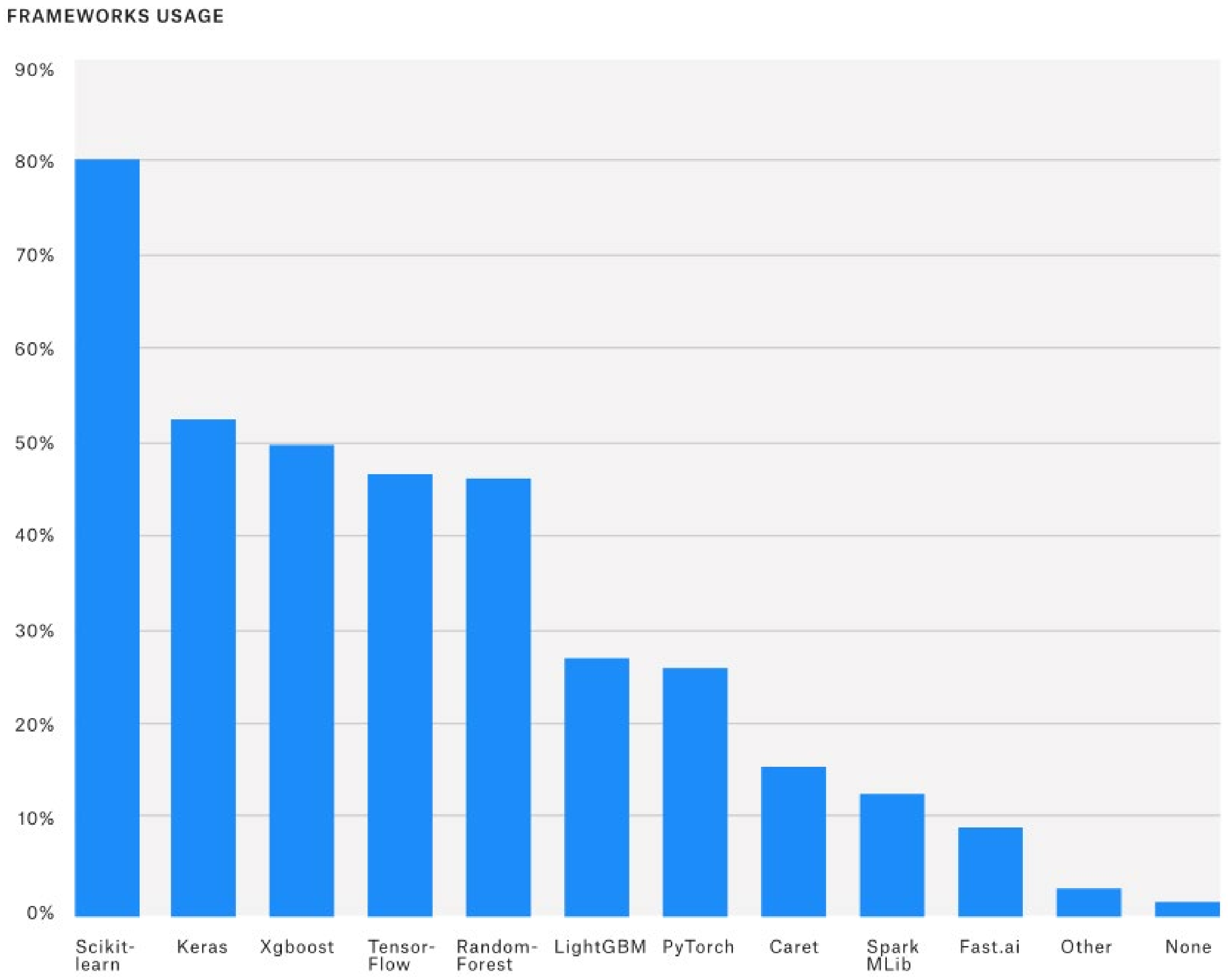
Kaggle: Making Data Science a Sport

https://www.kaggle.com/kaggle-survey-2019



Kaggle: Making Data Science a Sport

https://www.kaggle.com/kaggle-survey-2019



Tools This Semester

We will use python notebooks and popular data science libraries

Option 1)

Install Anaconda

From Anaconda install JupyterLab

Option 2)

Google Colab

We will start coding NEXT CLASS. You must come prepared!

First Resources

You should also check out these python resources:

A Whirlwind Tour of Python by Jake VanderPlas.

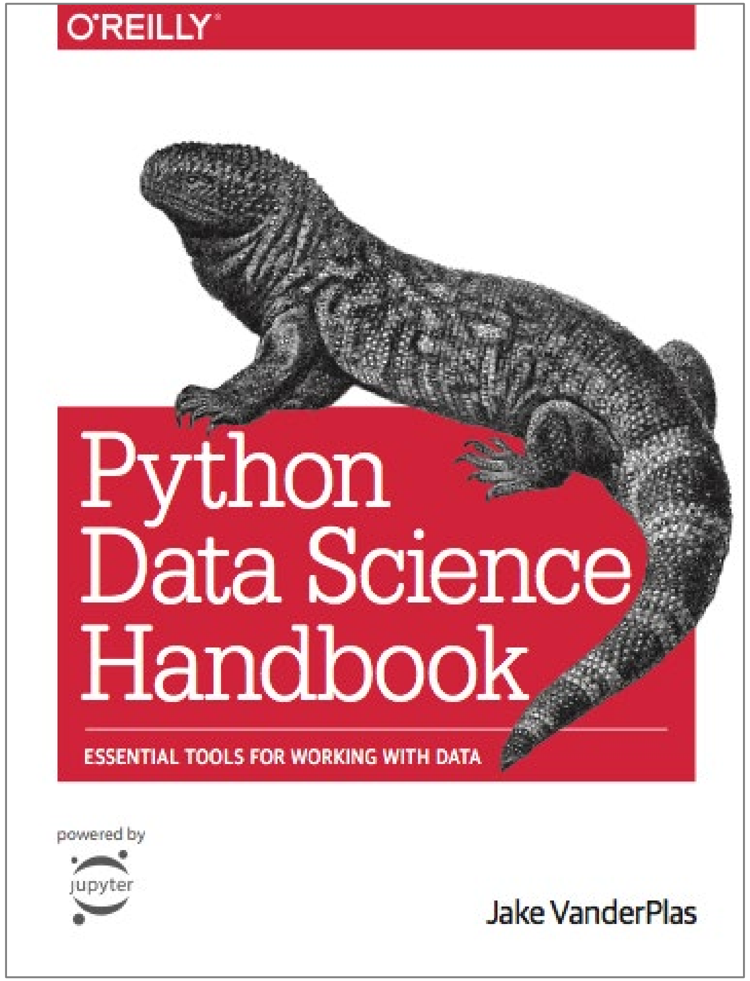
Freely available online at:

https://jakevdp.github.io/WhirlwindTourOfPython/

Python Data Science Handbook by Jake VanderPlas.

Freely available online at:

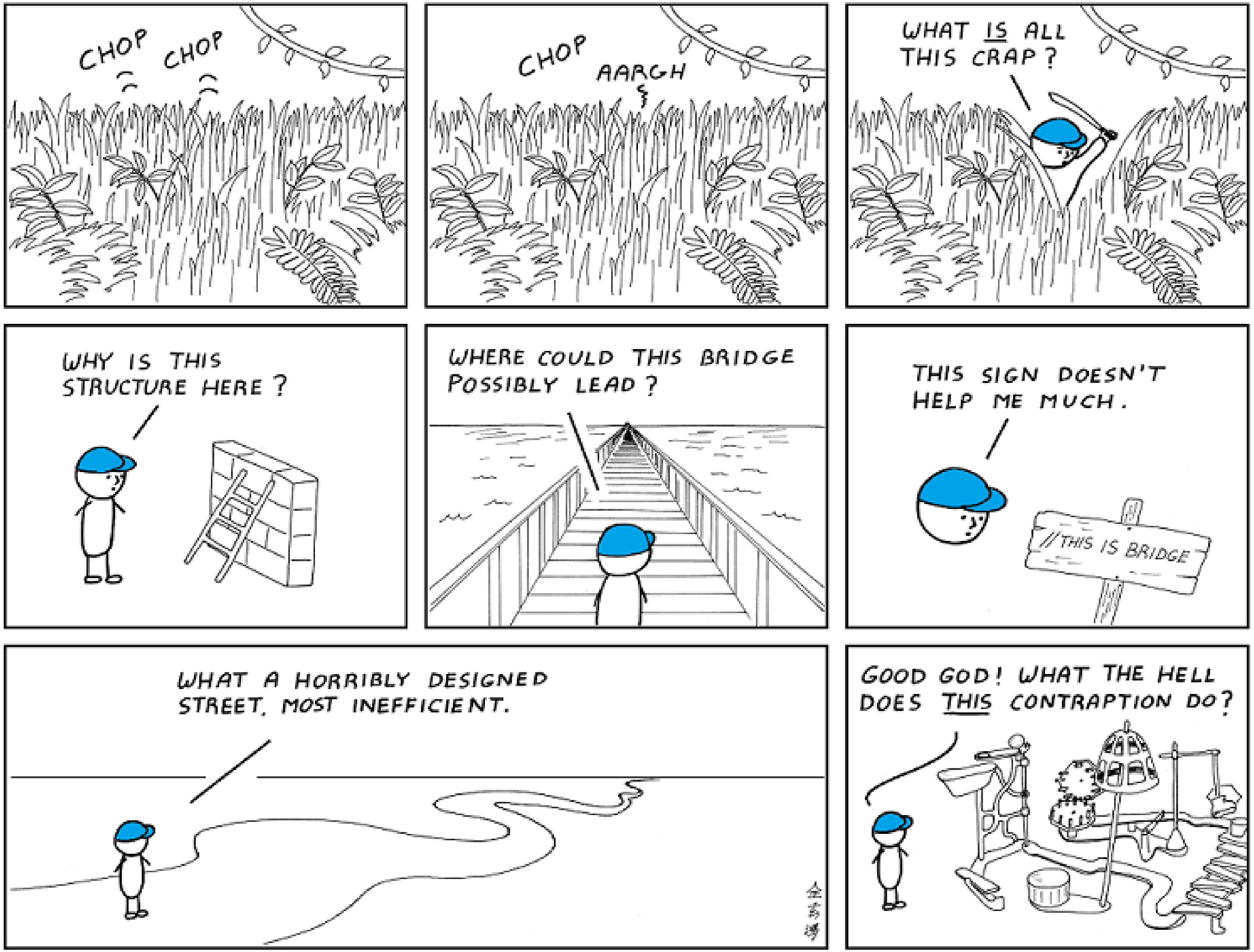
https://jakevdp.github.io/PythonDataScienceHandbook/



Part 3

CLASS INFO

Syllabus



I hate reading other people's code.