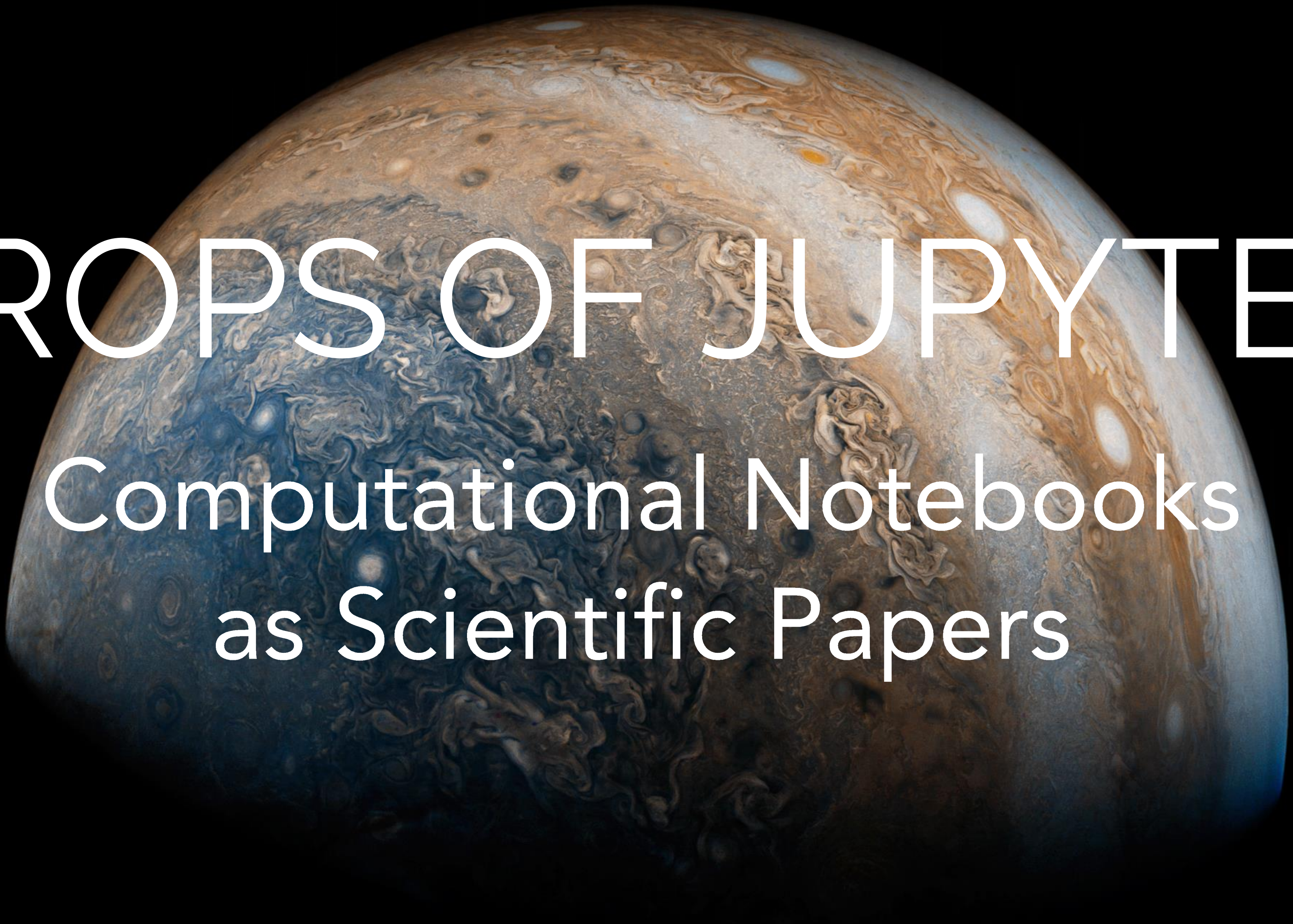


DROPS OF JUPYTER

Computational Notebooks as Scientific Papers

Stefan Uddenberg

Quick poll…

How I used to do my research…

How I used to do my research…

- Produce many different things

How I used to do my research…

- Produce many different things
  - E.g., notes about ideas, stimulus parameters, analysis scripts, figures with many variants…

How I used to do my research...

- Produce many different things
  - E.g., notes about ideas, stimulus parameters, analysis scripts, figures with many variants...
- Bring it all together in a slide deck, or ‘notebook’

How I used to do my research…

- Produce many different things
  - E.g., notes about ideas, stimulus parameters, analysis scripts, figures with many variants…
- Bring it all together in a slide deck, or ‘notebook’ — **MANUALLY**

How I used to do my research…

- Produce many different things
  - E.g., notes about ideas, stimulus parameters, analysis scripts, figures with many variants…
- Bring it all together in a slide deck, or ‘notebook’ — **MANUALLY**
- Write a separate paper/**manuscript**

Revealing Mental Defaults with Serial Reproduction

Stefan Uddenberg
03-05-2015

Introduction

When you picture a person walking down the street, what kind of person is it?

The goal of this project is to characterize the nature of such default representations through the method of serial reproduction.

Hypotheses

- Magnet hypothesis
  - Reproductions will be attracted to the far ends of the spectrum, likely in a 50/50 ratio. That is, the first deviation is up to chance, and errors simply snowball from there on out.
- Identity hypothesis
  - Your default face looks like you. For example, white subjects should have white default faces.
- Bayesian hypothesis
  - Your default face is a representation of the most likely face you will encounter. This means that most US subjects' default face should be white.

Experiment 1a – Default Race

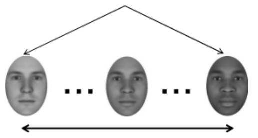
Participants

- Subjects were recruited via Amazon Mechanical Turk and paid $0.30 for their time.
  - Subjects were barred from using Google Chrome and mobile devices (e.g. tablet computers, mobile phones). Any attempt to access the website with a noncompliant device or browser was met with a fatal error message, and as such they were not allowed to complete the study.
- Collected 30 chains of 10 white US-based subjects each (N = 300). Originally aimed for 20 chains only.
- Segregated chains based on race.

Experiment 1 – Default Race

Stimuli

Created anchor faces with FaceGen



61 faces

Made intermediate faces with FantaMorph

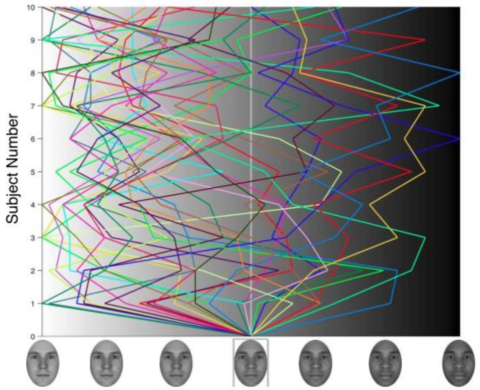
Used the SHINE toolbox to match the luminance

Experiment 1e – Luminance Matched Race

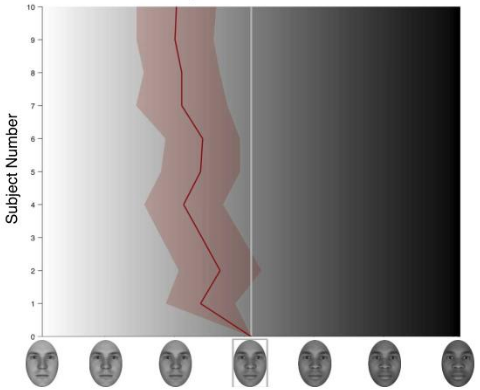
Method

- Pre-task
  - After accepting the advertised HIT, participants navigated to our website where they were greeted with a boilerplate consent form.
  - They then entered their demographic information on this page (race, gender, age, worker ID, and nationality), and then clicked the "Consent/Next" button and performed a basic browser compatibility check. Importantly, repeat subjects (as determined by their worker IDs) were barred from participating again.
  - On passing the browser compatibility test, they moved on to the task.
- Task
  - Each of the 30 chains started with the face in the midpoint of the continuum (namely face 31).
  - Subjects were first shown a face (202 px X 284 px) presented in Gaussian white noise (mean = 0, variance = 0.05) for 1 second, followed by a blank screen for 1 second. They then tried to reproduce the first face to the best of their ability by morphing an on-screen test face using a wheel slider (diameter = 580 px). The first test face shown varied at random, as did its corresponding position on the wheel slider. As subjects moved the mouse around the circle, the face smoothly transitioned from white to black and back again. To make their final choice, they merely clicked anywhere else on-screen and hit the "Submit" button.
  - The next subject was shown the previous subject's choice, but presented in noise.
  - Attention checks and exclusion criteria
    - Before the task began there was a countdown from 5 to 0, where one of the numbers turned red. Subjects entered which number turned red at the end of the study (which was always either 1 or 2, unbeknownst to them). Their data were excluded if they got this answer wrong, or if they took longer than 60 seconds to reproduce the face.

White subjects – 30 chains



White subjects – mean chain



Results

1. How many points are white-of-center (i.e. left of the start point/the grey line) in the preceding mean graph?

***10 out of 10 ($z = 3.16$, $p = .002$)

Conclusion: There is a systematic bias whiteward.

Results

2. How many individual steps whiteward (i.e. left of each preceding step)?

5 out of 10 ($z = 0$, $p = 1.0$)

Conclusion: Equal numbers of jumps whiteward and blackward, so the whiteward jumps *must* be bigger.

Results

3. Which steps are significantly different from their preceding steps?

Parametrically (alpha = .005):

– None significant (all $t(29)$s < 2.98, ps $>= .0058$)

Nonparametrically (deviations white vs black):

– None significant ($p >= .043$)

Conclusion: Any effect Whiteward accrues over time.

Results

4. Which steps are significantly different from the original starting point?

Parametrically (Table 1a):

– 4, 7 – 10 (all $t(29)$s $>= 3.11$, ps $<= .0042$, ds $<=$ 1.15

Nonparametrically (Table 1b):

– 9 (24/30 Whiteward, $z = 3.29$, $p = .001$)

Conclusion: The bias whiteward is quite large and takes 4 – 7 steps to get all the way there.

Table 1a - Parametric

Comparing each step to chain starting point (i.e. Face #31)

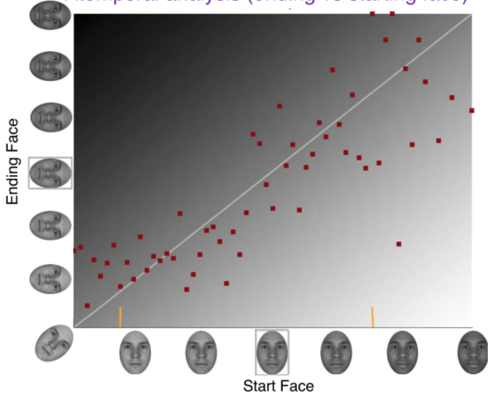
Step #	N	MD	SD	SEM	t	p
1	30	7.27	13.35	2.44	-2.98	0.006
2		4.47	15.90	2.90	-1.54	0.135
3		7.07	15.58	2.84	-2.48	0.019
4		9.70	15.16	2.77	-3.51	0.002
5		7.27	15.20	2.78	-2.62	0.014
6		6.97	14.29	2.61	-2.67	0.012
7		9.97	17.57	3.21	-3.11	0.004
8		9.97	14.61	2.67	-3.74	0.001
9		10.93	14.81	2.70	-4.04	0.000
10		10.73	15.31	2.80	-3.84	0.001

Table 1b - Nonparametric

Comparing each step to chain starting point (i.e. face #31)

Step #	N	N-White	N-Black	N-Stagnant	Z	p
1	30	21	9	0	-2.19	0.043
2		18	12	0	-1.10	0.362
3		19	10	1	-1.83	0.099
4		19	11	0	-1.46	0.200
5		20	10	0	-1.83	0.099
6		23	7	0	-2.92	0.005
7		22	8	0	-2.56	0.016
8		21	7	2	-2.92	0.005
9		24	6	0	-3.29	0.001
10		21	8	1	-2.56	0.016

Atemporal analysis (ending vs starting face)



Results

5. How many of the average points in the preceding graphs are in white space?

31 out of 57 (z = .662, p = .597, N = 57, N-Blacker = 25, N-whiter = 31, N-stagnant = 1)

Conclusion: Similar numbers of faces move both Whiter and Blacker.

Results

6. If we look only at the faces with some wiggle room, how many of those have averages in white space?

23 out of 39 (z = 1.30, p = .337, N = 39, N-White = 23, N-Black = 15, N-stagnant = 1)

Conclusion: No more faces in the middle range move white than black.

*"Wiggle room" meaning that we take only those faces whose value plus or minus their average deviation wouldn't put them outside the face space (faces 8 – 46 in this case).

Results

7. If we only look at the faces with some wiggle room*, are their deviations stat. sig.? Counting all faces.

Parametrically:

** $t(232) = 3.19$, $p = .002$, $d = .418$ (M = 2.373, SD = 11.37, SEM = .745)

Nonparametrically (# above vs. below line if shown on dev. graph):

** 93 out of 233 move Black ($z = 3.08$, $p = .003$, N = 233, N-White = 133, N-Black = 93, N-Stagnant = 7)

Conclusion: The middle faces overwhelmingly veer Whiteward.

*"Wiggle room" meaning that we take only those faces whose value plus or minus their average deviation wouldn't put them outside the face space (faces 8 – 46 in this case).

Results

8. How many of the whitest faces' ending points are in white space?

** 3 out of 18 (*z* = 2.83, *p* = .008, N = 18, N-White = 3, N-Black = 14, N-Stagnant = 1)

Conclusion: The Whitest faces veer Blackward.

Using faces 1 – 7 in this case.

Results

9. If we only look at the Whitest faces are their deviations stat. sig.? Counting all faces.

Parametrically:

- *** $t(46) = 6.41$, $p = 7.05*10^{-8}$ (M = -9.74, SD = 10.42, SEM = 1.52)

Nonparametrically (# above vs. below line if shown on dev. graph):

- *** 5 out of 47 (*z* = 5.40, *p* = $2.46*10^{-8}$, N = 47, N-White = 5, N-Black = 41, N-Stagnant = 1)

Conclusion: The Whitest faces overwhelmingly veer Blackward.

*Using faces 1 – 7 in this case.

Results

10. How many of the blackest faces' ending points are in white space?

8 out of 11 (*z* = 1.51, *p* = .227, N = 11, N-White = 8, N-Black = 3, N-Stagnant = 0)

Conclusion: The Blackest faces trend Whiteward.

Using faces 47 – 61 in this case.

Results

11. If we only look at the blackest faces are their deviations stat. sig.? Counting all faces.

Parametrically:

- ** $t(19) = 3.77$, $p = .001$ (M = 11.35, SD = 13.47, SEM = 3.01)

Nonparametrically (# above vs. below line if shown on dev. graph):

- * 15 out of 20 (*z* = 2.24, *p* = .041, N = 20, N-White = 15, N-Black = 5, N-Stagnant = 0)

Using faces 47 – 61 in this case.

Conclusions

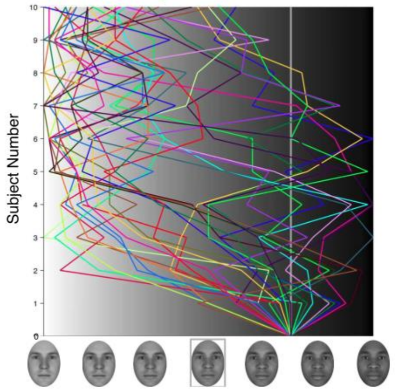
- Subjects spend more time in white space for those faces in the middle where they are free to go in either direction, and make bigger jumps toward white-space than toward black-space.
- Subjects veer blackward when trying to reproduce the whitest faces, and blackward when trying to reproduce the whitest faces.
- These data are consistent with both the identity and weighted-average hypotheses, but not the magnet hypothesis. We can tease the two apart by collecting more data for minority participants.
- Effects are a bit weaker when you take away differences in overall luminance (White faces look like they are in darker lighting, Black faces look washed out).

Experiment 1f – LM Race Start Black

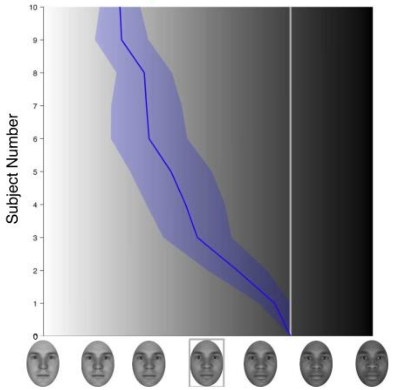
Method

- All methods, stimuli and relevant details are the same as Experiment 1e except as noted below:
  - Started black (face #46/61) instead of in the middle.

White subjects – 30 chains



White subjects – mean chain



Results

1. How many points are white-of-center (i.e. left of the start point/the grey line) in the preceding mean graph?

***10 out of 10 ($z = 3.16$, $p = .002$)

Conclusion: There is a systematic bias whiteward.

Results

2. How many individual steps whiteward (i.e. left of each preceding step)?

***10 out of 10 ($z = 3.16$, $p = .002$)

Conclusion: Whiteward ever, blackward never.

Results

3. Which steps are significantly different from their preceding steps?

Parametrically (alpha = .005):

- *Step 3 [$t(29) = 3.341$, $p = .002$]

Nonparametrically (deviations white vs black):

- ~None significant ($p >= .0052$ – Step 2)

Conclusion: Individual jumps are small, so the effect Whiteward accrues over time.

Results

4. Which steps are significantly different from the original starting point?

Parametrically (Table 1a):

- 2 – 10 (all $t(29)$s >= 3.67, ps <= .00095, ds <= 1.36

Nonparametrically (Table 1b):

- 3, 5 – 10 ($p <= .001$)

Conclusion: The bias whiteward is quite large and takes 4 – 7 steps to get all the way there.

Table 1a - Parametric

Comparing each step to chain starting point (i.e. Face #46)

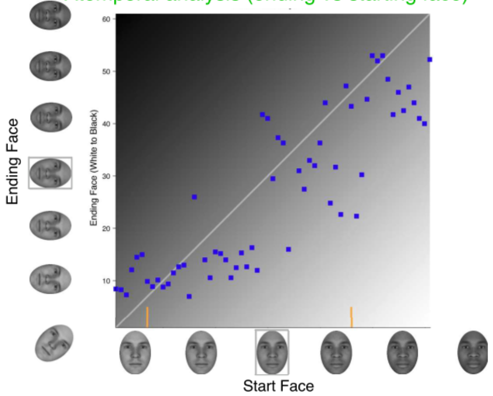
Step #	N	MD	SD	SEM	t	p
1	30	2.87	7.68	1.40	-2.05	0.050
2		9.70	14.45	2.64	-3.68	0.001
3		16.97	16.63	3.04	-5.59	0.000
4		19.10	19.16	3.50	-5.46	0.000
5		21.77	19.93	3.64	-5.98	0.000
6		25.77	18.67	3.41	-7.56	0.000
7		26.23	17.22	3.14	-8.34	0.000
8		26.63	13.50	2.47	-10.80	0.000
9		30.77	13.06	2.38	-12.91	0.000
10		31.07	9.92	1.81	-17.15	0.000

Table 1b - Nonparametric

Comparing each step to chain starting point (i.e. face #46)

Step #	N	N-White	N-Black	N-Stagnant	Z	p
1	30	20	9	1	-2.19	0.043
2		23	7	0	-2.92	0.005
3		25	5	0	-3.65	0.000
4		23	7	0	-2.92	0.005
5		24	6	0	-3.29	0.001
6		25	4	1	-4.02	0.000
7		25	5	0	-3.65	0.000
8		29	1	0	-5.11	0.000
9		30	0	0	-5.48	0.000
10		30	0	0	-5.48	0.000

Atemporal analysis (ending vs starting face)



Results

5. How many of the average points in the preceding graphs are in white space?

** 40 out of 59 (z = 2.734, p = .0086, N = 59, N-Blacker = 19, N-whiter = 40, N-stagnant = 0)

Conclusion: More faces end up whiter than they started (on average).

Results

6. If we look only at the faces with some wiggle room, how many of those have averages in white space?

** 26 out of 35 (z = 2.449, p = .0060, N = 35, N-White = 26, N-Black = 9, N-stagnant = 0)

Conclusion: Those faces with wiggle room move whiter on average.

*"Wiggle room" meaning that we take only those faces whose value plus or minus their average deviation wouldn't put them outside the face space (faces 7 – 43 in this case).

Results

7. If we only look at the faces with some wiggle room*, are their deviations stat. sig.? Counting all faces.

Parametrically:

** $t(167) = 4.371$, $p = 2.17*10^{-5}$, d = .337 (M = 4.083, SD = 12.109, SEM = .934)

Nonparametrically (# above vs. below line if shown on dev. graph):

** 48 out of 167 move Black (z = 5.555, $p = 1.66 * 10^{-8}$, N = 168, N-White = 110, N-Black = 48, N-Stagnant = 10)

Conclusion: The middle faces overwhelmingly veer Whiteward.

*"Wiggle room" meaning that we take only those faces whose value plus or minus their average deviation wouldn't put them outside the face space (faces 7 – 43 in this case).

Results

8. How many of the whitest faces' ending points are in white space?

** 0 out of 6 (z = 2.449, p = .031, N = 6, N-White = 0, N-Black = 6, N-Stagnant = 0)

Conclusion: The Whitest faces veer Blackward.

Using faces 1 – 6 in this case.

Results

9. If we only look at the Whitest faces are their deviations stat. sig.? Counting all faces.

Parametrically:

- *** $t(47) = 5.669$, $p = 8.49 \times 10^{-7}$, $d = .818$ (M = -7.104, SD = 8.682, SEM = 1.253)

Nonparametrically (# above vs. below line if shown on dev. graph):

- *** 11 out of 48 go White (z = 3.752, $p = 2.22 \times 10^{-4}$, N = 48, N-White = 11, N-Black = 36, N-Stagnant = 1)

Conclusion: The Whitest faces overwhelmingly veer Blackward.

*Using faces 1 – 6 in this case.

Results

10. How many of the blackest faces' ending points are in white space?

*** 16 out of 18 (z = 3.30, $p = 7.63 \times 10^{-6}$, N = 18, N-White = 16, N-Black = 0, N-Stagnant = 2)

Conclusion: The Blackest faces go Whiteward.

Using faces 44 – 61 in this case.

Results

11. If we only look at the blackest faces are their deviations stat. sig.? Counting all faces.

Parametrically:

** $t(83) = 5.523$, $p = 3.74 \times 10^{-7}$, $d = .603$ (M = 6.988, SD = 11.597, SEM = 1.265)

Nonparametrically (# above vs. below line if shown on dev. graph):

- *** 20 out of 84 go Black(z = 4.801, $p = 1.58 \times 10^{-6}$, N = 84, N-White = 63, N-Black = 20, N-Stagnant = 1)

Using faces 44– 61 in this case.

Conclusions

- We replicated the results of Expt. 1e despite the fact that we started so far into black space.
- This suggests that Griffiths' hypothesis is right: it doesn't matter where you start off with, because the outcome will always converge on the prior.

Alternative possibility…

- Attraction to middle of spectrum
  - In all the other studies we have run (age, shapes, gender, emotion) none of the chains have a bias away from the middle, and starting on the far ends inexorably leads to reproductions veering middle-ward.
  - This raises the question of whether our earlier race expt. results are simply due to subjects' bias to where they *think* the middle of the spectrum is.
  - This experiment directly tests subjects' ability to pinpoint the middle of the spectrum, even when only exposed to some subset thereof.

Experiment 1g – Perceptual Middle

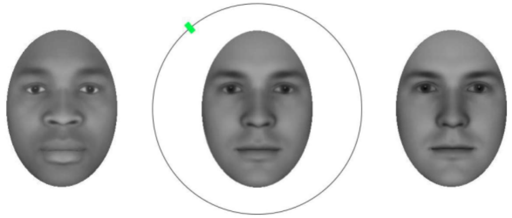
Participants

- Subjects were recruited via Amazon Mechanical Turk and paid $0.30 for their time.
  - Subjects were barred from using Google Chrome and mobile devices (e.g. tablet computers, mobile phones). Any attempt to access the website with a noncompliant device or browser was met with a fatal error message, and as such they were not allowed to complete the study.
- Collected 100 US-based subjects (N =100).
- Universal chains, not segregated on the basis of race.

Experiment 1g – Perceptual Middle

Task

What is the face exactly halfway between Faces 1 & 2?



Experiment 1g – Perceptual Middle

Method

- Pre-task
  - After accepting the advertised HIT, participants navigated to our website where they were greeted with a boilerplate consent form.
  - They then entered their demographic information on this page (race, gender, age, worker ID, and nationality), and then clicked the "Consent/Next" button and performed a basic browser compatibility check. Importantly, repeat subjects (as determined by their worker IDs) were barred from participating again.
  - On passing the browser compatibility test, they moved on to the task.
- Task
  - Subjects were initially shown the two anchor faces (202 px X 284 px) in the spectrum at the top of the screen (either white on the left or black on the left at random), and told they would need to identify the face that was exactly halfway between/a perfect blend of the two faces.
  - On clicking "Start", the instructions disappeared and they were confronted with the two anchor faces flanking the circle. They were given the standard countdown attention check, and then shown a face chosen at random from the distribution of faces (its corresponding position on the wheel slider also varied randomly).
  - However, unlike in previous studies, they were not shown the whole continuum of faces. Half the subjects had their continuum anchored 5 faces from the whitest face, while the other half's were anchored 5 faces away from the blackest face (faces 6 and 56 respectively). The other end point of their spectrum could then be a face chosen at random between the other anchor and 5 faces away from the mid-point of the full spectrum on the non-anchored side (faces 36 and 26 respectively). For example, a subject whose spectrum was anchored on white could end up seeing faces 6 to 36 – 56.
  - They then tried to produce the absolute middle face to the best of their ability by morphing an on-screen test face using a wheel slider (diameter = 580 px). As subjects moved the mouse around the circle, the face smoothly transitioned from white to black and back again. To make their final choice, they merely clicked anywhere else on-screen and hit the "Submit" button.
  - Attention checks and exclusion criteria
    - Before the task began there was a countdown from 5 to 0, where one of the numbers turned red. Subjects entered which number turned red at the end of the study (which was always either 1 or 2, unbeknownst to them). Their data were excluded if they got this answer wrong, or if they took longer than 60 seconds to reproduce the face.

Results

Results

1. Is the distribution of choices made in this experiment different from the true middle of the spectrum?

***$t(99) = 3.654$, $p = .0004$, $d = .734$ (M = 29.00, SD = 5.47, SEM = .547)

** 66/100 go white-of-center, $p = .0018$(N-White = 66, N-Black = 30, N-Stagnant = 4)

Conclusion: Yes, there is a strongly statistically significant bias whiter, but it is very small in absolute terms (2/61 faces = 3.28%).

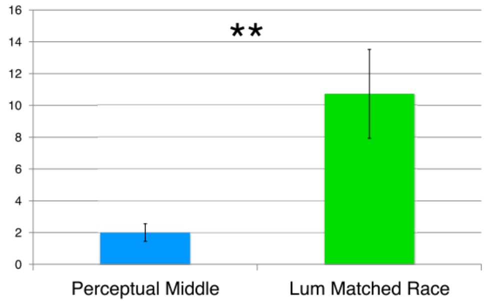
Results

2. Is the distribution of choices made in this experiment different from the distribution of *endpoints* (N=30) from the serial luminance-matched experiment?

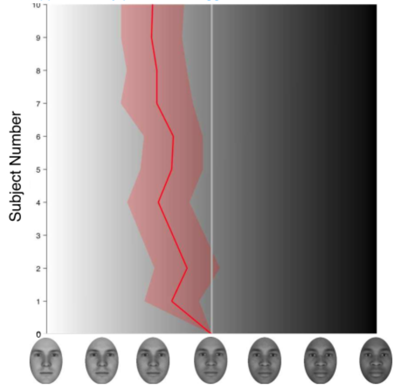
***$t(31.253) = 3.066$, $p = .0044$, $d = .734$ (MD = 8.733, SED = 2.848)

Conclusion: Yes, such that the serial endpoints are much whiter. These results remain exactly the same whether you use only the first 30 points from this study to make the Ns equal, and even if you use the last 3 steps from the serial experiment (N=90).

Perceptual middle vs Lum Matched Race.
Error bars are ±1 SEM.



Mean chain from Luminance Matched Race vs. current result (approx). Blue box is approximately placed and bigger than the confidence interval.



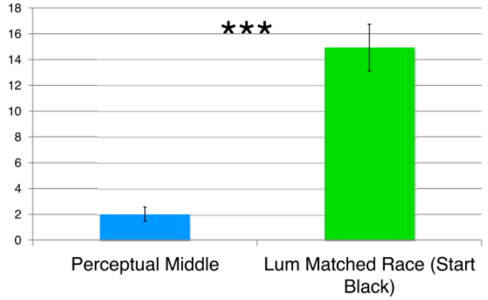
Results

3. Is the distribution of choices made in this experiment different from the distribution of *endpoints* (N=30) from the serial luminance-matched experiment *when starting black*?

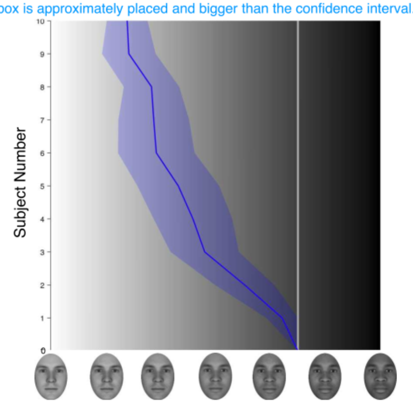
***$t(34.451) = 7.432$, $p = 1.18*10^{-8}$, $d = 2.532$ (MD = -14.0667, SED = 1.893)

Conclusion: Yes, such that the serial endpoints are much whiter. These results remain exactly the same whether you use only the first 30 points from this study to make the Ns equal, and even if you use the last 3 steps from the serial experiment (N=90).

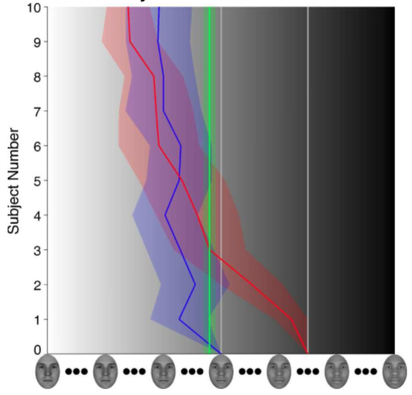
Perceptual middle vs Start Black.
Error bars are ±1 SEM.



Mean chain from Luminance Matched Race vs. current result (approx). Blue box is approximately placed and bigger than the confidence interval.



White subjects – both chains + PM



Conclusions

- The perceptual middle is very slightly biased white, but nowhere near enough to explain our results.

Experiment 1h – Just Noticeable Race

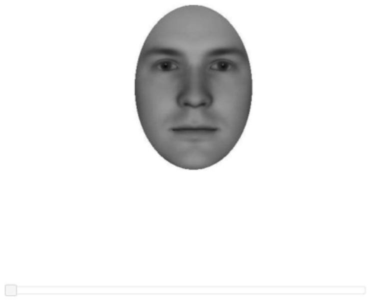
Participants

- Subjects were recruited via Amazon Mechanical Turk and paid $0.30 for their time.
  - Subjects were barred from using mobile devices (e.g. tablet computers, mobile phones). Any attempt to access the website with a noncompliant device or browser was met with a fatal error message, and as such they were not allowed to complete the study.
- Collected 76 – 88 US-based subjects (N =164). Still collecting data as of 03/05/2015.
- Chains were segregated on the basis of race; data shown are from white subjects.

Experiment 1h – Just Noticeable Race

Task

What is the first face that looks black/white?



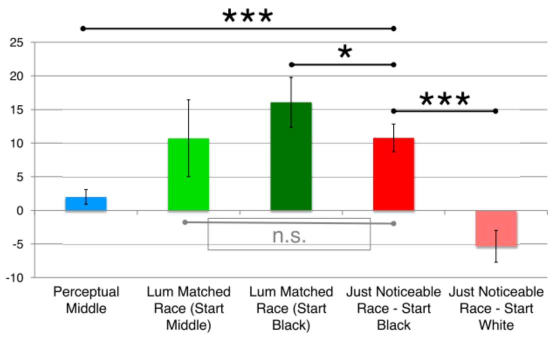
Experiment 1h – Just Noticeable Race

Method

- Pre-task
  - After accepting the advertised HIT, participants navigated to our website where they were greeted with a boilerplate consent form.
  - They then entered their demographic information on this page (race, gender, age, worker ID, and nationality), and then clicked the "Consent/Next" button and performed a basic browser compatibility check. Importantly, repeat subjects (as determined by their worker IDs) were barred from participating again.
  - No mandatory browser compatibility test was included in this version.
- Task
  - Subjects were initially shown a single anchor face (202 px X 284 px) from the endpoints of the spectrum at the top of the screen (either white or black at random), and told they would need to move the slider slowly to the right in order to identify the first face they found that was definitely a member of the other race.
  - On clicking "Start", the instructions disappeared and they were confronted with the anchor faces over a simple slider (549 px wide, because it's divisible by 61), which always started all the way to the left. Once they moved the slider to the right the face transformed from white to black (or vice versa). After moving the slider and 5 seconds elapsed, they were allowed to move on when they found the face they wanted and pressed the "Next" button that appeared.
  - Attention checks and exclusion criteria
    - Subjects were excluded if they chose the beginning face, or if they got the attention check wrong ("What is four plus three?").

Results

Perceptual middle vs Lum Matched Race vs Just Noticeable Race. Error bars are 95% CI.



Results Summary

- When starting White and looking for a Black face, subjects choose ≈face 20, which is the end-point of the original Start Middle experiment and statistically indistinguishable from it. It is, however, significantly different from the end-point of the Start Black, and from the perceptual middle.
- When starting Black and looking for a White face, everything comes out significant at $p < .001$ ($2.5 * 10^{-5}$, in fact.)
- Black space starts a lot closer to the middle (≈face 36) than does white space (≈face 20)
  - $t(154.4) = 3.478$, $p = .0006$. Arrived at by looking at absolute difference between the two (so 10.8 vs 5.33)

Conclusions

- Subjects are going *at least as white* as the first noticeably white face on average in the serial reproduction studies.

Experiment 1i – Exploring scale space

Participants

- Subjects were recruited via Amazon Mechanical Turk and paid $0.30 for their time.
  - Subjects were barred from using mobile devices (e.g. tablet computers, mobile phones). Any attempt to access the website with a noncompliant device or browser was met with a fatal error message, and as such they were not allowed to complete the study.
- Collected 23 & 26 chains for weighted black and weighted white respectively. Collecting 20 x 30 replications as of 03-05-2015.
- Chains were segregated on the basis of race; data shown are from white subjects.

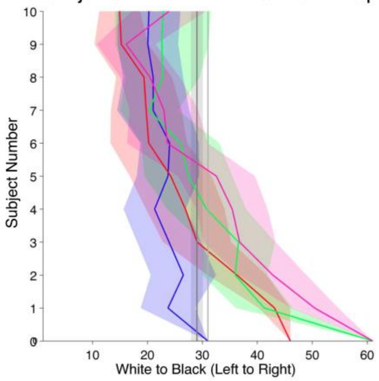
Experiment 1i – Exploring scale space

Method

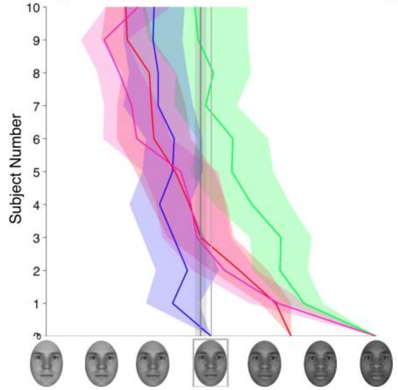
- Stimuli
  - We used two 61 face-long continua identical in their end-points, but differing in their scales. One was weighted white (with 45 faces in white space) and the other weighted black (with 45 faces in black space).
- Pre-task
  - After accepting the advertised HIT, participants navigated to our website where they were greeted with a boilerplate consent form.
  - They then entered their demographic information on this page (race, gender, age, worker ID, and nationality), and then clicked the "Consent/Next" button and performed a basic browser compatibility check. Importantly, repeat subjects (as determined by their worker IDs) were barred from participating again.
  - On passing the browser compatibility test, they moved on to the task.
- Task
  - Each of the 23/26 chains started with the very Blackest face (namely face 61).
  - Subjects were first shown a face (202 px X 284 px) for 1 second, followed by a blank screen for 1 second. They then tried to reproduce the first face to the best of their ability by morphing an on-screen test face using a wheel slider (diameter = 580 px). The first test face shown varied at random, as did its corresponding position on the wheel slider. As subjects moved the mouse around the circle, the face smoothly transitioned from white to black and back again. To make their final choice, they merely clicked anywhere else on-screen and hit the "Submit" button.
  - The next subject was shown the previous subject's choice, but presented in noise.
  - Attention checks and exclusion criteria
    - Before the task began there was a countdown from 5 to 0, where one of the numbers turned red. Subjects entered which number turned red at the end of the study (which was always either 1 or 2, unbeknownst to them). Their data were excluded if they got this answer wrong, or if they took longer than 60 seconds to reproduce the face, or if they noticed the weirdness of the scale.

Results

White subjects – mean chains in scale space



White subjects – mean chains in face/race space

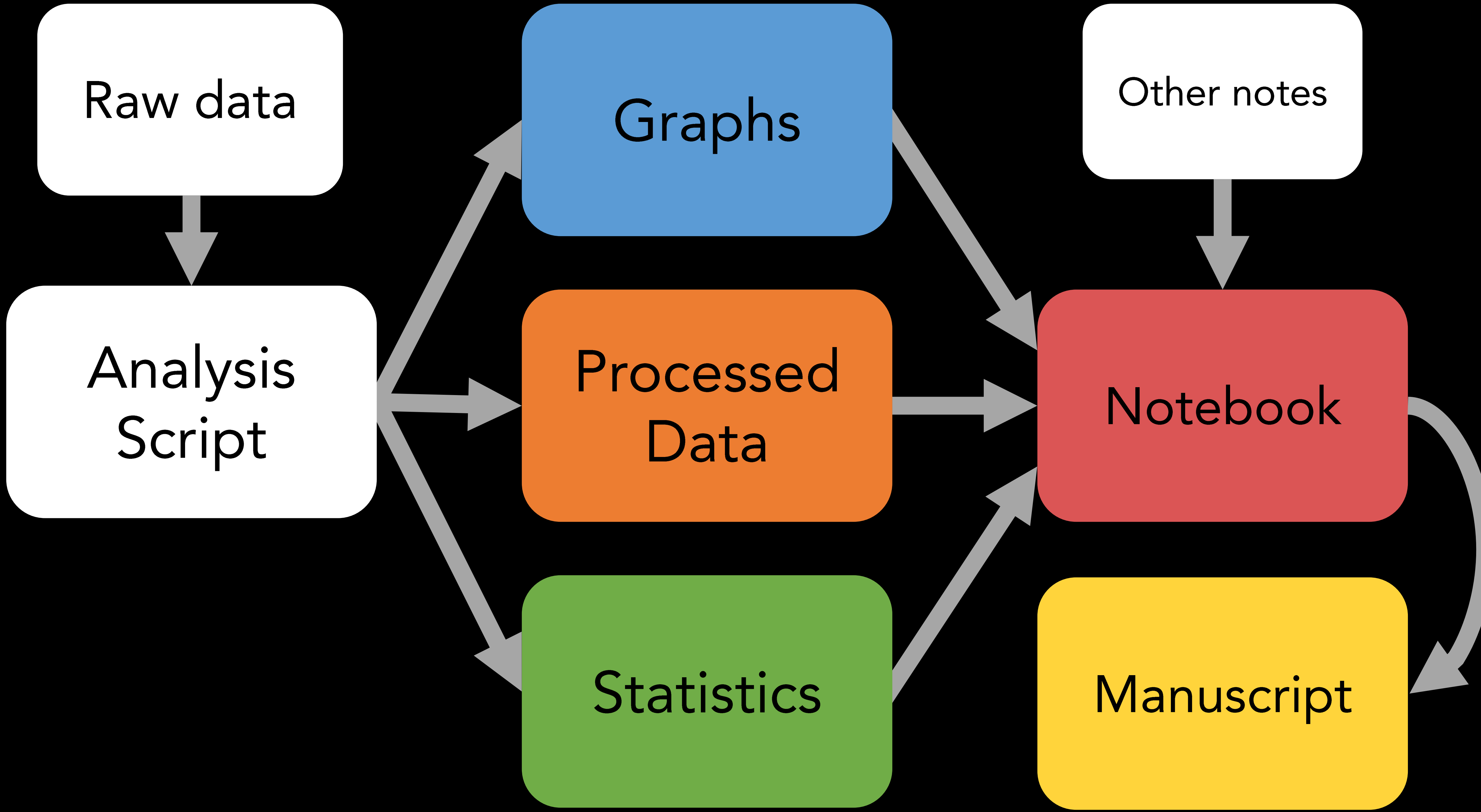


Results Summary

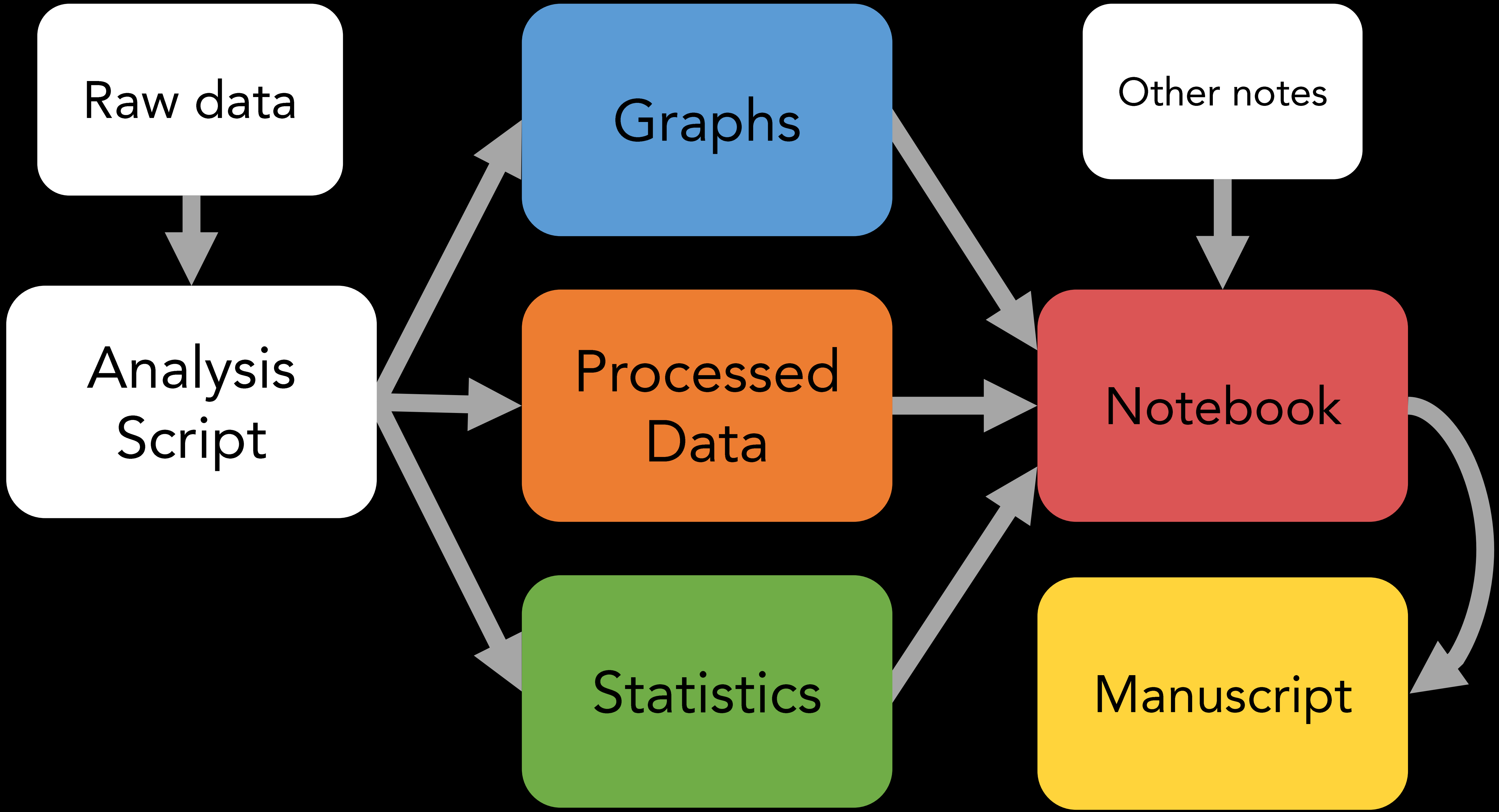
- This is worrying, because our results are consistent with the interpretation that they are just going to a particular point in the scale and that is has nothing to do with race (although why they are going to *that* particular point is still unknown).

Some updates from one project...

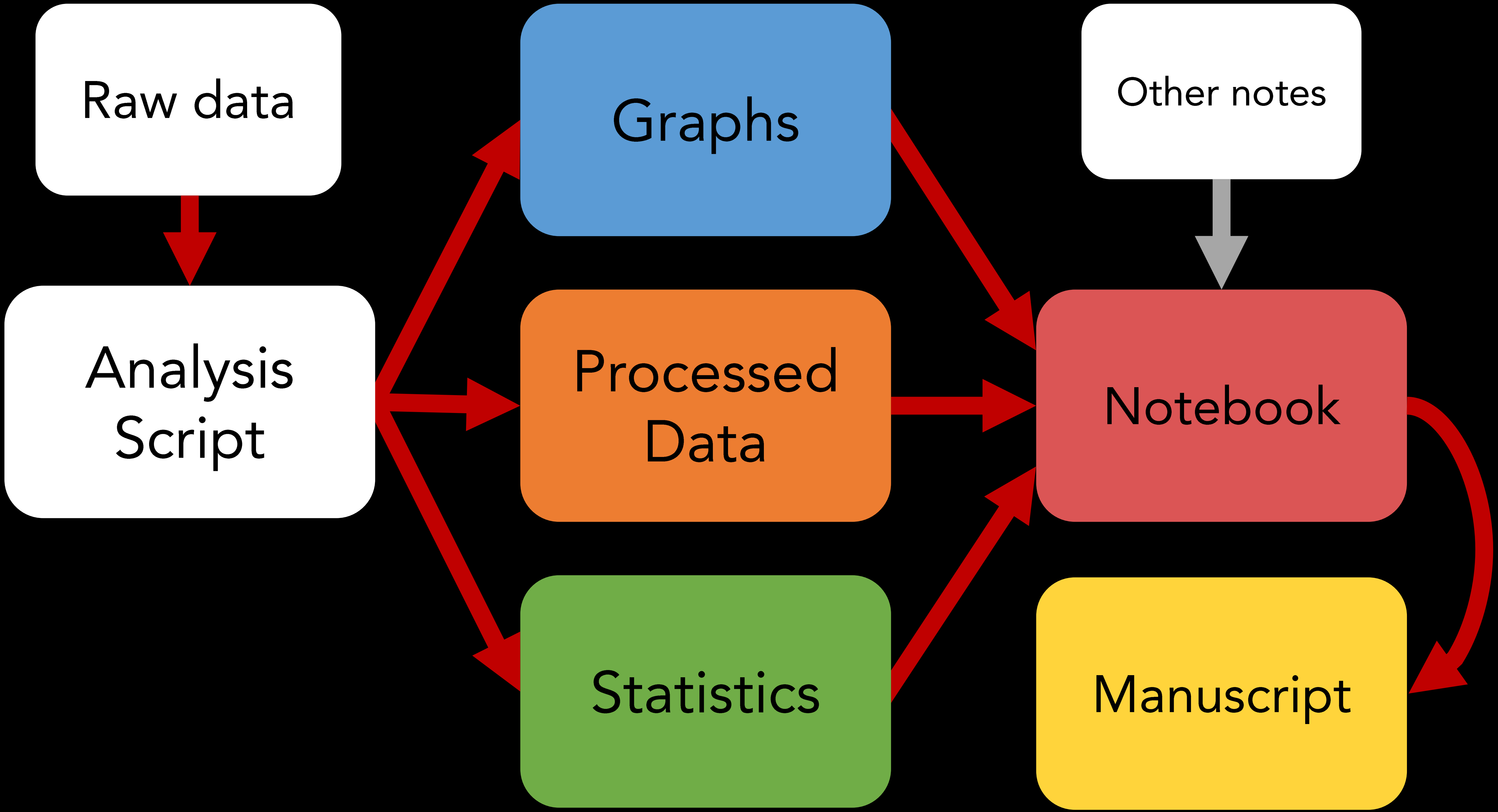
#	Name	Ext	Size	Type
6	Serial Reproduction Lab Notebook 05-29-2014.pptx	p...	13,967 ...	Microsoft Pow...
7	Serial Reproduction Lab Notebook 06-01-2014.pptx	p...	14,332 ...	Microsoft Pow...
8	Serial Reproduction Lab Notebook 07-30-2014.pptx	p...	25,639 ...	Microsoft Pow...
9	Serial Reproduction Lab Notebook PERCEPTUAL MIDDLE 07-21-2014.pptx	p...	453 KB	Microsoft Pow...
10	Serial Reproduction Lab Notebook PERCEPTUAL MIDDLE 09-05-2014.pptx	p...	453 KB	Microsoft Pow...
11	Serial Reproduction Lab Notebook PERCEPTUAL MIDDLE 09-28-2014.pptx	p...	451 KB	Microsoft Pow...
12	Serial Reproduction Lab Notebook RACE 03-05-2015.pptx	p...	6,666 KB	Microsoft Pow...
13	Serial Reproduction Lab Notebook RACE 04-18-2015.pptx	p...	6,666 KB	Microsoft Pow...
14	Serial Reproduction Lab Notebook RACE 04-25-2015.pptx	p...	272 KB	Microsoft Pow...
15	Serial Reproduction Lab Notebook RACE 06-03-2015.pptx	p...	263 KB	Microsoft Pow...
16	Serial Reproduction Lab Notebook RACE 07-30-2014.pptx	p...	6,344 KB	Microsoft Pow...


Raw data
Analysis Script
Graphs
Processed Data
Statistics
Other notes
Notebook
Manuscript

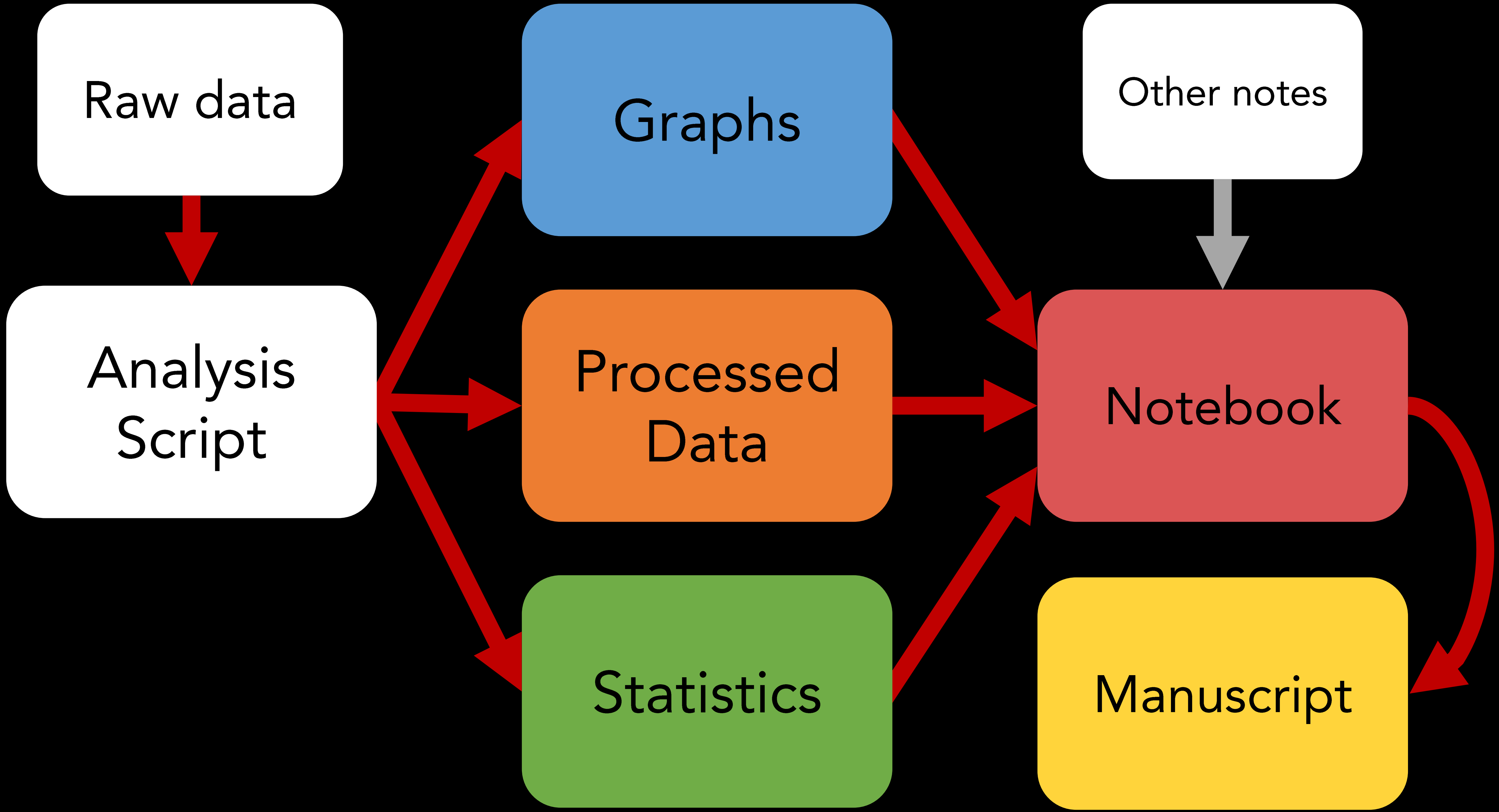
What happens when new data come in?



Raw data
Analysis Script
Graphs
Processed Data
Statistics
Other notes
Notebook
Manuscript


Raw data
Analysis Script
Graphs
Processed Data
Statistics
Other notes
Notebook
Manuscript

Make a notebook <u>once</u>



Raw data
Analysis Script
Graphs
Processed Data
Statistics
Other notes
Notebook
Manuscript

Computational Notebook

Prose

Code

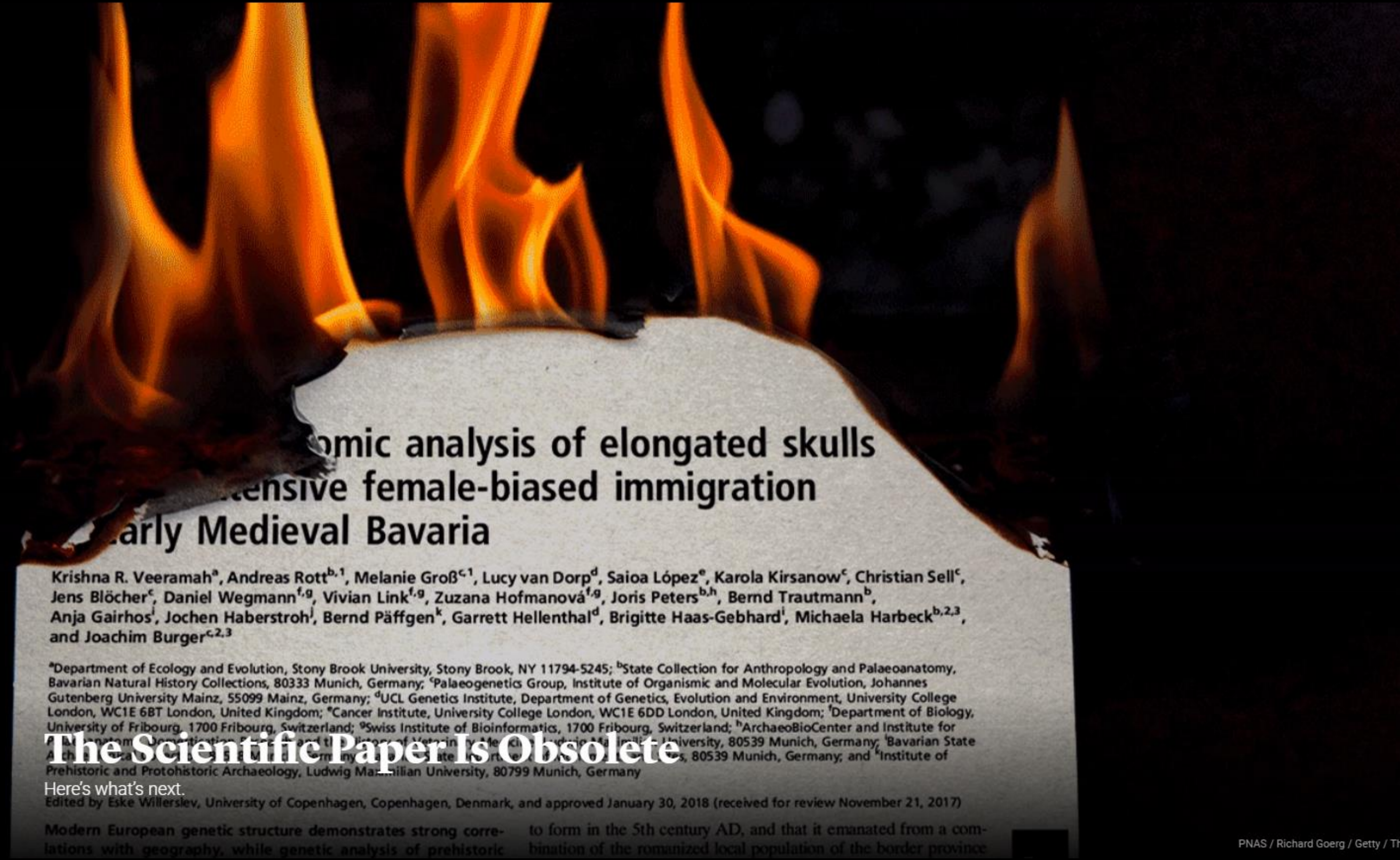
Demos

Graphs

Processed Data

Statistics

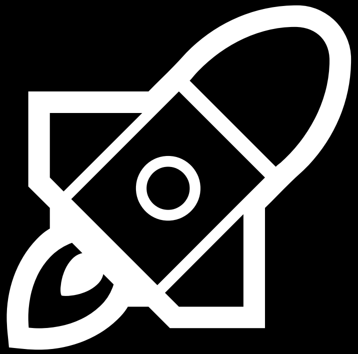
One project, one notebook



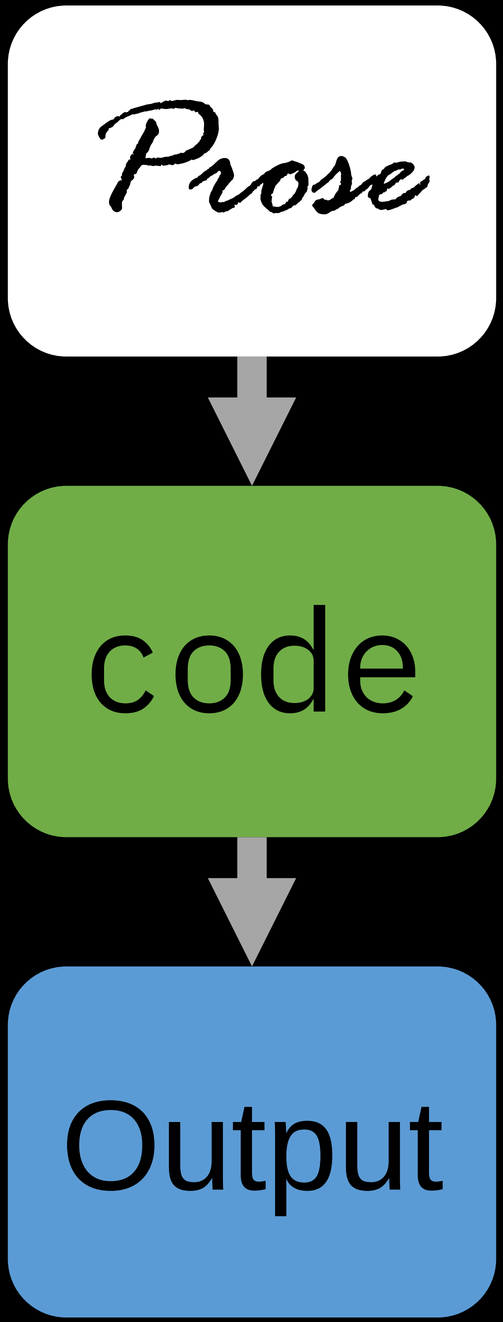
The Scientific Paper Is Obsolete

Here's what's next.

PNAS / Richard Goerg / Getty / Th

code + *prose* = 

‘Literate programming’



Computational Manuscript

Prose

Code

Demos

Graphs

Processed Data

Statistics

Computational notebooks can become **manuscripts**

Generate stats directly from analyses

$t(242)=1.39$, $p=.166$, $d=.178$

Generate stats directly from analyses

$t(242)=1.39$, $p=.166$, $d=.178$

Update numbers/figures in your papers **automatically**!

The Speed of Demography in Face Perception

Clara Colombatto, Stefan Uddenberg, and Brian J. Scholl

Yale University

Running head: Speed of Demography

Address for correspondence:
Brian J. Scholl
Department of Psychology
Yale University
Box 208205
New Haven, CT 06520-8205
brian.scholl@yale.edu
203-432-4629

Word count: XXX (Main text + abstract)

Version: Test version – not for submission

January 31, 2019

2 Introduction

If any old picture is worth a thousand words, then pictures of faces may be worth considerably more than that: when viewing faces, we form impressions of a wide variety of properties, even beyond the extraction of identity information (for a review see Todorov, 2017). Some of these properties reflect relatively stable traits — such as how trustworthy, competent, or dominant a person is. Other properties reflect relatively transient states — as when a person looks momentarily focused, distracted, or disappointed. And still other properties reflect our personal experience or preferences — as when faces strike us as familiar, or as attractive. The fact that we so readily 'read' faces in this way seems extraordinary given that all faces share the same basic features and overall configuration, and given that so many of our impressions of faces seem so ineffable.

2.1 Perceiving Demography?

Some of the most intriguing traits that we discern from faces, however, go beyond those already mentioned: in addition, we also readily perceive demographic characteristics, such as race, age, and gender. Such demographic features are undeniably important, given how they influence both other psychological processes (such as memory; e.g. Meissner & Brigham, 2001) and real-world outcomes (as when a person's race influences police officers' split-second decisions about whether to shoot a potential perpetrator; e.g. Correll et al., 2007). But there is some disagreement about just how fundamental demographic properties are in social perception.

In some ways, properties such as age, race, and gender seem even more foundational than other features. For example, whereas it is controversial whether facial expressions of emotion are universal (e.g. Gendron et al., 2014), every society has both men and women. And whereas there is considerable debate about the degree to which split-second inferences about social traits like trustworthiness are reliable signals (e.g. Todorov et al., 2015), characteristics such as perceived age are especially honest cues. (Indeed, aging is associated with a number of structural and textural changes in faces [e.g. Berry & McArthur, 1986; George & Hole, 2000] that are notoriously difficult to conceal, though not for any lack of trying by the cosmetic industry.) And more generally, demographic traits often lead us to categorize others on the basis of social groups (such as race; e.g. Cloutier et al., 2005), which does not happen so readily on the basis of social traits such as extraversion.

At the same time, there are other indications that at least some demographic traits may not be so foundational in social perception after all. For example, race and gender seem to radically diverge in terms of how irresistibly they are encoded into memory (Cosmides et al., 2003). In "who said what?" memory tests, for example, errors rarely if ever cross gender boundaries regardless of whether gender is made salient or not. But for race the situation is more complicated: while early studies suggested that race was similarly automatically encoded, later studies revealed that such memory traces largely vanish when there are other competing 'coalitional' cues (such as shirt color; Kurzban et al., 2001; Pietraszewski et al., 2015). These results suggest that demographic cues such as race may not be so primitive after all — such that 'race can be erased' (while gender cannot).

2.2 The Current Studies: Seeing, Fast and Slow

The current experiments seek to contribute to the larger project of determining how foundational demographic properties are in social perception, using an especially crude but apt metric: speed. Studies of the speed of face perception are legion. And the bottom line is that face perception, in general, is fast. For example, past work has identified the amount of time it takes to detect the presence of a human face among distractors (~ 240-290 ms; Rousselet et al., 2003), to detect the presence of a familiar face (~ 360-470 ms; Barragan-Jason et al., 2012; Besson et al., 2012), or to recognize the identity of a particular individual

2 Introduction

If any old picture is worth a thousand words, then pictures of faces may be worth considerably more than that: when viewing faces, we form impressions of a wide variety of properties, even beyond the extraction of identity information (for a review see Todorov, 2017). Some of these properties reflect relatively stable traits — such as how trustworthy, competent, or dominant a person is. Other properties reflect relatively transient states — as when a person looks momentarily focused, distracted, or disappointed. And still other properties reflect our personal experience or preferences — as when faces strike us as familiar, or as attractive. The fact that we so readily 'read' faces in this way seems extraordinary given that all faces share the same basic features and overall configuration, and given that so many of our impressions of faces seem so ineffable.

2.1 Perceiving Demography?

Some of the most intriguing traits that we discern from faces, however, go beyond those already mentioned: in addition, we also readily perceive demographic characteristics, such as race, age, and gender. Such demographic features are undeniably important, given how they influence both other psychological processes (such as memory; e.g. Meissner & Brigham, 2001) and real-world outcomes (as when a person's race influences police officers' split-second decisions about whether to shoot a potential perpetrator; e.g. Correll et al., 2007). But there is some disagreement about just how fundamental demographic properties are in social perception.

In some ways, properties such as age, race, and gender seem even more foundational than other features. For example, whereas it is controversial whether facial expressions of emotion are universal (e.g. Gendron et al., 2014), every society has both men and women. And whereas there is considerable debate about the degree to which split-second inferences about social traits like trustworthiness are reliable signals (e.g. Todorov et al., 2015), characteristics such as perceived age are especially honest cues. (Indeed, aging is associated with a number of structural and textural changes in faces [e.g. Berry & McArthur, 1986; George & Hole, 2000] that are notoriously difficult to conceal, though not for any lack of trying by the cosmetic industry.) And more generally, demographic traits often lead us to categorize others on the basis of social groups (such as race; e.g. Cloutier et al., 2005), which does not happen so readily on the basis of social traits such as extraversion.

At the same time, there are other indications that at least some demographic traits may not be so foundational in social perception after all. For example, race and gender seem to radically diverge in terms of how irresistibly they are encoded into memory (Cosmides et al., 2003). In "who said what?" memory tests, for example, errors rarely if ever cross gender boundaries regardless of whether gender is made salient or not. But for race the situation is more complicated: while early studies suggested that race was similarly automatically encoded, later studies revealed that such memory traces largely vanish when there are other competing 'coalitional' cues (such as shirt color; Kurzban et al., 2001; Pietraszewski et al., 2015). These results suggest that demographic cues such as race may not be so primitive after all — such that 'race can be erased' (while gender cannot).

2.2 The Current Studies: Seeing, Fast and Slow

The current experiments seek to contribute to the larger project of determining how foundational demographic properties are in social perception, using an especially crude but apt metric: speed. Studies of the speed of face perception are legion. And the bottom line is that face perception, in general, is fast. For example, past work has identified the amount of time it takes to detect the presence of a human face among distractors (~ 240-290 ms; Rousselet et al., 2003), to detect the presence of a familiar face (~ 360-470 ms; Barragan-Jason et al., 2012; Besson et al., 2012), or to recognize the identity of a particular individual

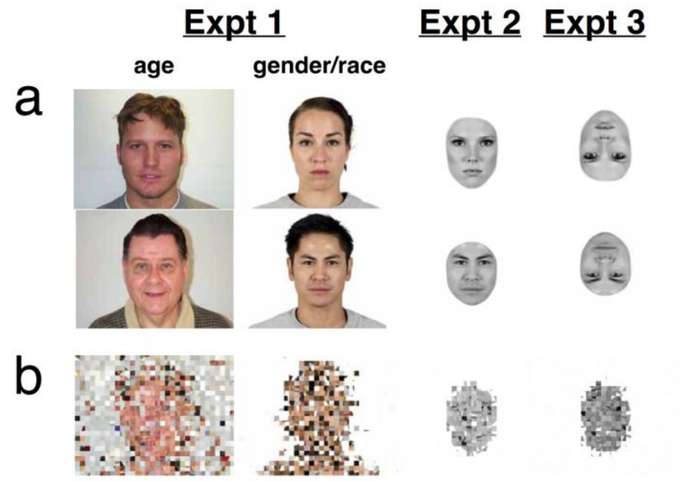


Figure 3.1: A depiction of (a) the face stimuli and (b) the masks used throughout the experiments.

For each of these three sets of faces separately, 120 masks were created by adjoining several ($0.68° \times 0.68°$) tiles, each taken from a randomly chosen picture within the same set; the positions of all colored tiles (excluding the white background) were scrambled, and the tiles were rotated in one of four orientations (0°, 90°, 180°, 270°; see Fig. 1b for example masks).

3.1.4 Procedure

Each trial began with a black fixation cross ($0.74° \times 0.71°$) on a white background presented in the center of the display for 500ms, followed by the face stimulus ($22.75° \times 16.10°$ in the gender and race tasks, and $21.35° \times 16.10°$ for the age task). On unspeeded blocks, the face was surrounded by a green frame (0.17° wide) and remained visible until response. On speeded blocks instead, the face was surrounded by a red frame and remained visible for a limited exposure time, after which it was replaced by a mask (of the same size, and also surrounded by a red frame) appearing for 500ms, and then by a blank screen with a green frame until response.

Observers pressed a key from 1 to 5 on a standard keyboard to indicate their response where 1 and 5 indicated the extremes ("Definitely Female" and "Definitely Male" in the gender task, "Definitely White" and "Definitely Black" in the race task, and "Definitely Young" and "Definitely Old" in the age task), and 3 indicated "Not sure". Reminders for the response mapping were present throughout the whole trial below the bounding box in Helvetica font: the numerals 1 through 5 (0.68° tall, highest number 9.40° below the center of the display; first number 10.30° to the left of the center of the display), with their labels jupyter

Goals of this tutorial

- Define modeling and understand why models are useful.
- Learn about different strategies for fitting models.
- Learn how to fit a simple model to empirical data using an iterative least-squares routine.
- Learn about model selection and learn how to use cross-validation for model selection.

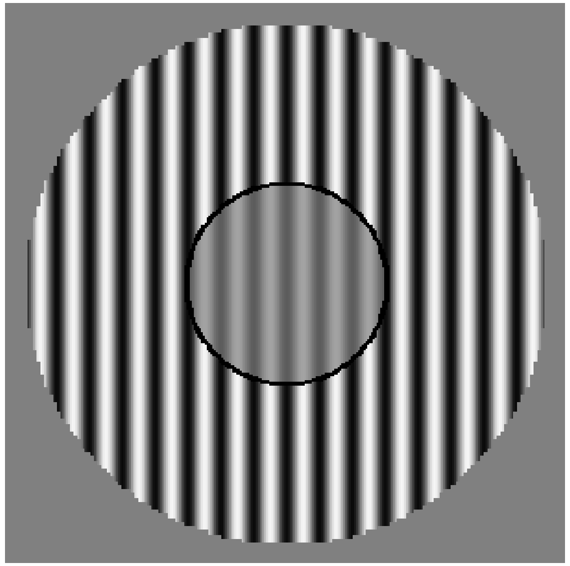
Why model?

Let's consider a data-set taken from an experiment in visual neuroscience. In this experiment, participants viewed a stimulus that contained a grating. In each trial two gratings were displayed. To show these stimuli, let's import some stuff from IPython:

In [1]:
```
from IPython.display import display, Image
```

The first grating that was shown was surrounded by another grating:

In [2]:
```
display(Image(filename='images/surrounded.png'))
```



The second grating was not surrounded:

A world of **possibilities**

Turn a **notebook**
into a **presentation**

This works for live websites, too!

Reproduce analyses from **any web browser**

binder

Create (and auto-grade) homework assignments with `nbgrader`

https://www.youtube.com/watch?v=5WUm0QuJdFw

Make a literal book with `jupyter_book`

Be **flexible**

Jupyter

Julia

Python

R

120 kernels to choose from...

...and no need to choose!

Computational notebooks can make your work more…

Computational notebooks can make your work more…

- Organized

Computational notebooks can make your work more…

- Organized
- Comprehensible

Computational notebooks can make your work more…

- Organized
- Comprehensible
- Reproducible

Computational notebooks can make your work more…

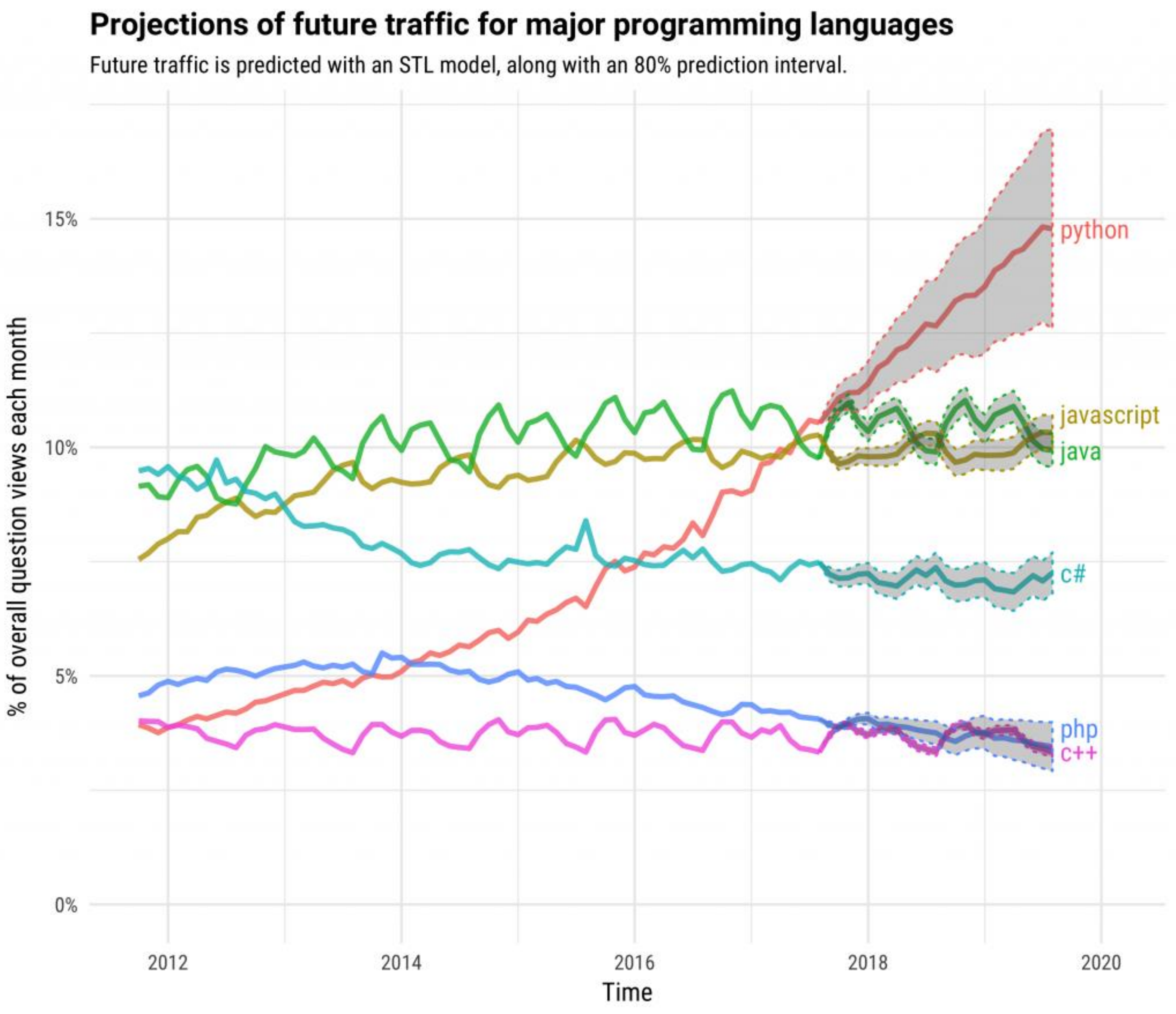
- Organized
- Comprehensible
- Reproducible
- Shareable (e.g., HTML, PDF, LaTeX)

Computational notebooks can make your work more…

- Organized
- Comprehensible
- Reproducible
- Shareable (e.g., HTML, PDF, LaTeX)
- Flexible: write in multiple languages!

Of all possible languages, today's tutorial will be in **Python**

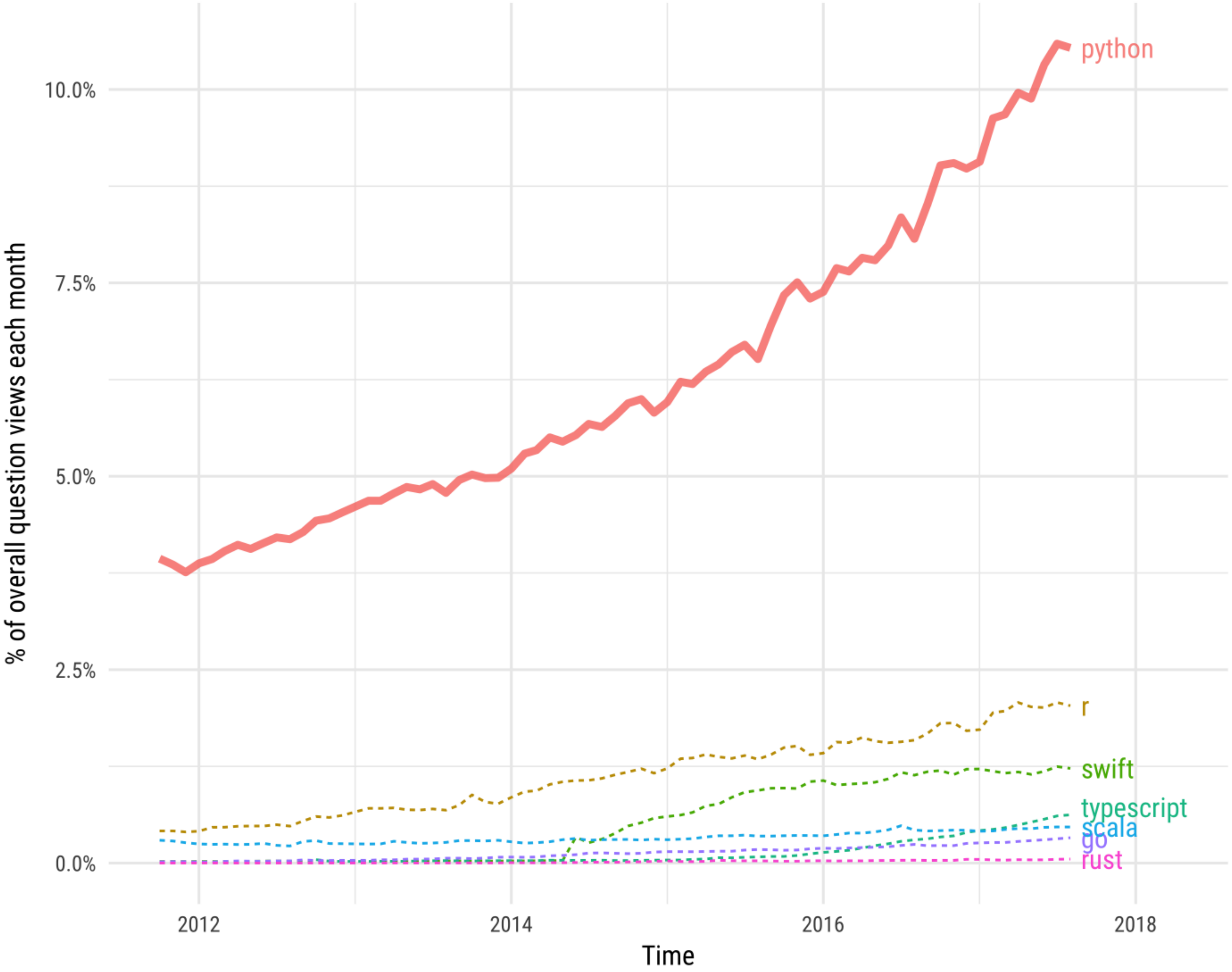
Python is **big** and **growing**



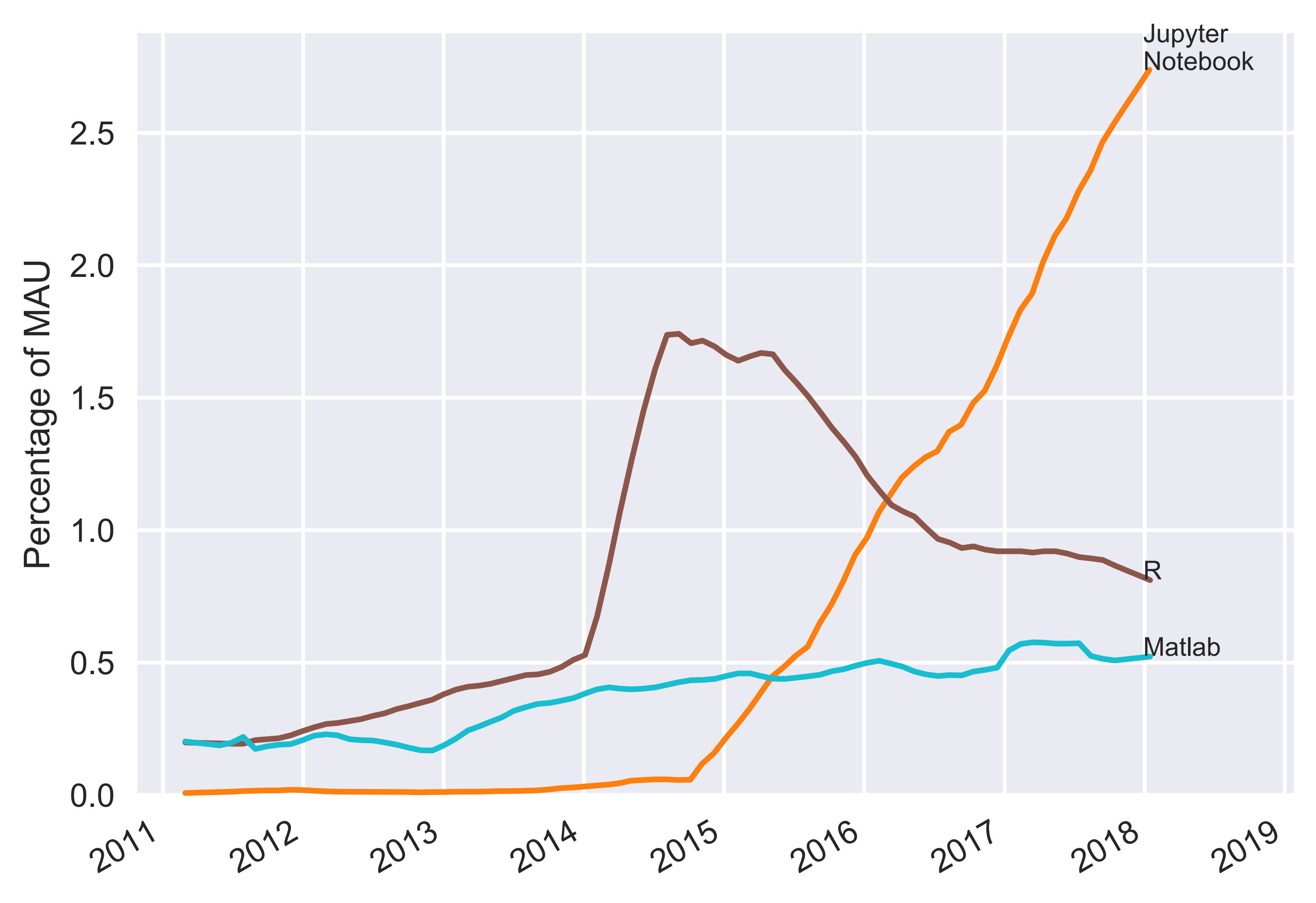
Projections of future traffic for major programming languages
Future traffic is predicted with an STL model, along with an 80% prediction interval.
% of overall question views each month
15%
10%
5%
0%
2012
2014
2016
2018
2020
Time
python
javascript
java
c#
php
c++

Python compared to smaller, growing technologies

Based on question traffic in World Bank high-income countries



Jupyter alone has more users than R



"You can't just copy-pase pseudocode into a program and expect it to work"

Always use the best tool for the job…

Always use the best tool for the job…

But Python is usually
at least the 2nd best tool…

Always use the best tool for the job...

But Python is usually
at least the 2nd best tool...

And almost always the
easiest tool of all!

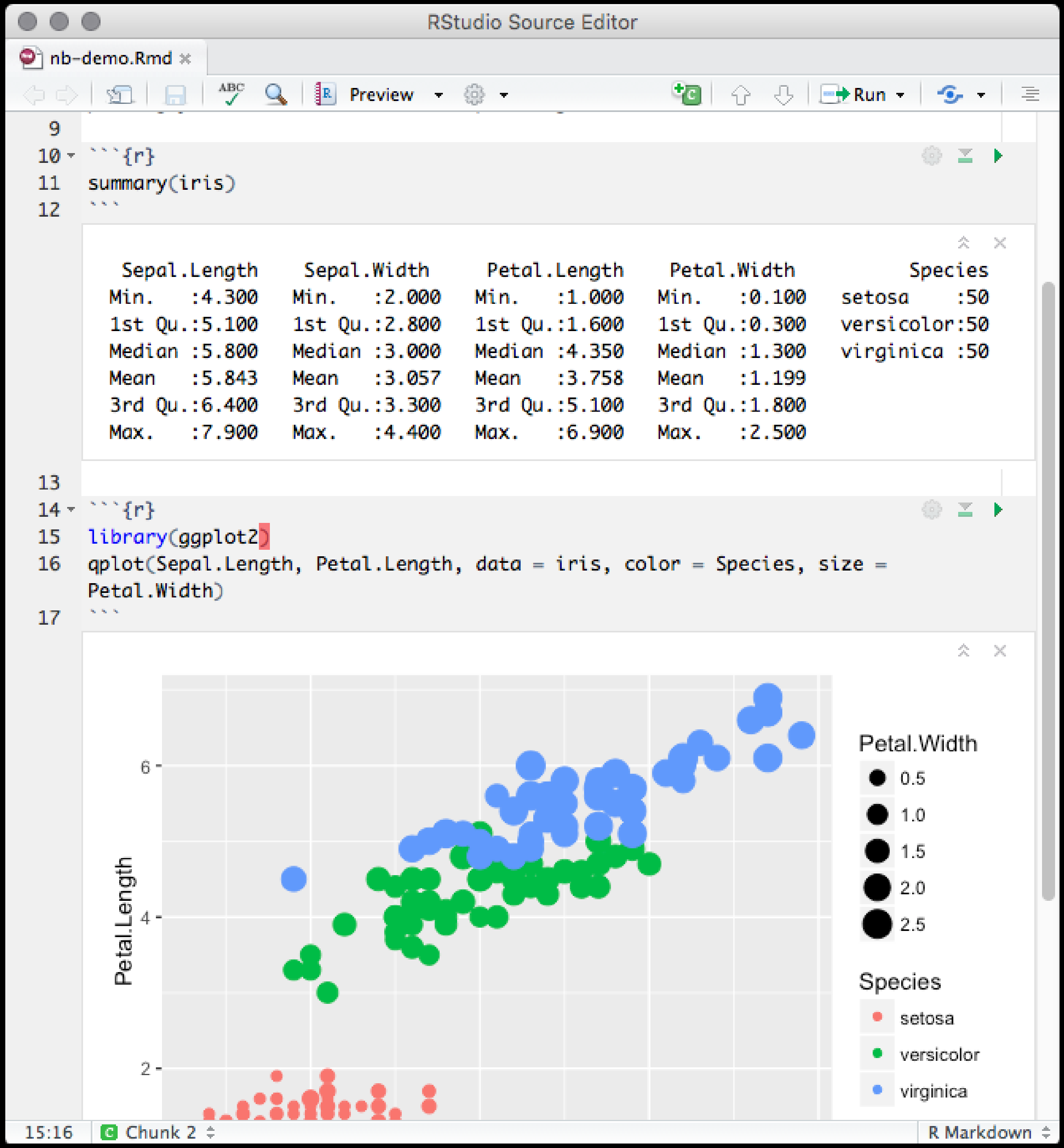
...and the best tool probably still works with Jupyter

BUT!

If you like **R**, you could also go with **R Markdown**.

Rstudio is amazing, and you can make manuscripts with packages like **`papaja`**, **`tufte`**, and **`rticles`**

https://libscie.github.io/rmarkdown-workshop/handout.html



Let's dive in

1. Jupyter notes: https://goo.gl/UYkJDk
2. PsiPyPublish: https://goo.gl/zJXRu7