

#SASGF

VIRTUAL

SAS® GLOBAL FORUM 2021

AMERICAS | MAY 18 - 20

ASIA PACIFIC | MAY 19 - 20

EMEA | MAY 25 - 26

It's In My DNA: A SAS® Program To Assist With Finding Unknown Relatives

Rick Langston

Rick Langston, before his retirement, was a senior developer and manager within the Platform R&D Division at SAS Institute. Rick's responsibilities included PROC FORMAT, functions, date/time processing, licensing software, and SAS/TOOLKIT Software. He has delivered keynotes on various aspects of SAS for many regional and local SAS user group conferences since 1994. Rick is a graduate of the University of North Carolina at Chapel Hill. He has been a SAS user since 1977, and worked at SAS Institute for over 38 years, from 1980 until December 2018. In his retirement, his hobbies include genealogy and DNA research.

Outline

- Quick intro to AncestryDNA and what I'm doing
- My read_matches4 SAS program to assist me
- Read_matches4 Part 1: reading the match data
- Read_matches4 Part 2: researching the relatives
- Highlights of the SAS code implementation
- Success stories

Outline Continued

- Problems with the non-SAS user community!
- Python as an implementation alternative
- Providing Python programs to users
- Comparing Python and SAS features



#SASGF

SAS® GLOBAL FORUM 2021

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Rick Langston's DNA Matches

List Map

Filter by:

Unviewed

Common ancestors

Messaged

Notes

Trees

Shared DNA

Groups

Search | Sort

Parent/Child



[Redacted Name]

Parent/Child

50% shared DNA | 3,456 cM



Unlinked tree

View match

...

Full Sibling



[Redacted Name]

Brother

Mother's Side

47% - 54% shared DNA | 2,639 cM



Linked tree | 63 People



Common ancestor

View match

...



[Redacted Name]

Brother

Mother's Side

42% - 49% shared DNA | 2,380 cM



Linked tree | 63 People



Common ancestor

View match

...

Close Family



1st – 2nd Cousin

Mother's Side

13% shared DNA | 904 cM ⓘ

 Unlinked tree

[View match](#)



1st – 2nd Cousin

Mother's Side

12% shared DNA | 821 cM ⓘ

 No family tree

[View match](#)



1st – 2nd Cousin

7% shared DNA | 497 cM ⓘ

 No family tree

[View match](#)



1st – 2nd Cousin

6% shared DNA | 445 cM ⓘ

 No family tree

[View match](#)



1st – 2nd Cousin

Mother's Side

6% shared DNA | 421 cM ⓘ

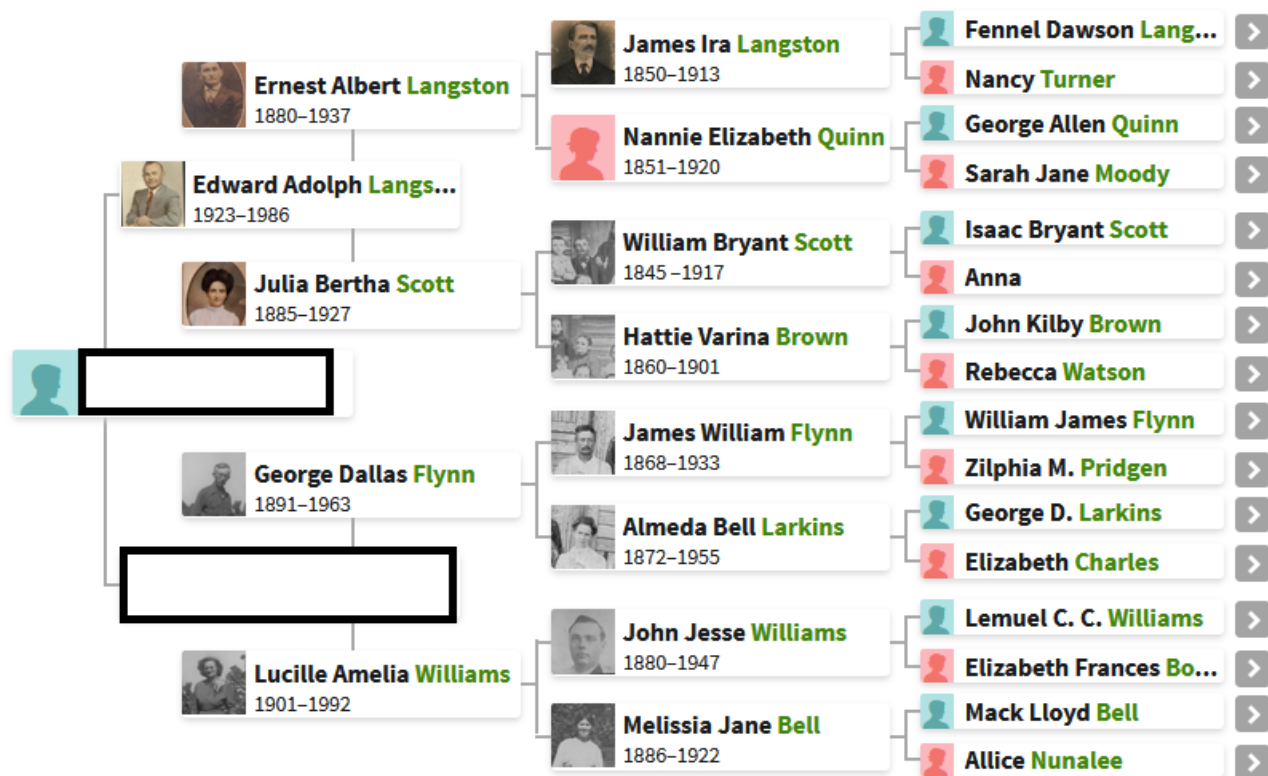
 No family tree

[View match](#)



's Linked Tree 63 People

This is a preview of the public tree linked to larry langston's DNA results. Surnames that appear in both your tree and larry langston's tree are **marked in green**.



What I'm Doing

- Helping people find unknown relatives
- Interpreting the DNA match results
- I've written the SAS program `read_matches4`
- It assists me in my searches
- Two parts to this program

Part 1: Reading the Match Data

- We need the HTML behind the scenes
- It can be captured with the Firefox browser
- The program will post-process the data
- We extract the pertinent parts
- We create a CSV

Rick Langston's DNA Matches

List Map

Filter by:

Unviewed

Common ancestors

Messed

Notes

Trees

Shared DNA

Groups

Search Sort

Parent/Child



- Open Link in New Tab
- Open Link in New Window
- Open Link in New Private Window

- Bookmark This Link

- Save Link As...

- Save Link to Pocket

- Copy Link Location

- Copy

- Search Google for "Rick Langston's..."

- Send Link to Device

- Print Selection

- View Selection Source

- Inspect Accessibility Properties

- Inspect Element (Q)



- Take a Screenshot



- Video DownloadHelper

A | 3,456 cM



Unlinked tree

View match

Full Sibling



ed DNA | 2,639 cM



Linked tree | 63 People



Common ancestor

View match



ed DNA | 2,380 cM



Linked tree | 63 People



Common ancestor

View match

Close Family



Unlinked tree

View match

```

1 <html class="mouseEvents" lang="en-US"><head><script type="text/javascript" async="" src="https://www.google.com/pagead/conversion_async
2 <script type="text/javascript">window.NREUM||(NREUM={});NREUM.info = {"agent":"","beacon":"bam-cell.nr-data.net","errorBeacon":"bam
3 <base href="https://www.ancestry.com/discoveryui-matches">
4 <link type="text/css" rel="stylesheet" href="https://www.ancestrycdn.com/ui/2.9.3/css/core.css">
5 <link type="text/css" rel="stylesheet" href="https://www.ancestrycdn.com/ui/2.9.3/css/guidance.css">
6 <link type="text/css" rel="stylesheet" href="https://www.ancestrycdn.com/ui/2.9.3/css/sidebar.css">
7
8 <link type="text/css" rel="stylesheet" href="https://www.ancestrycdn.com/shared-navigation/header/ancestry/2.1/content/header-289d2c
9 <meta name="viewport" content="width=device-width, initial-scale=1, user-scalable=no">
10 <link type="text/css" rel="stylesheet" href="https://www.ancestrycdn.com/shared-navigation/footer/ancestry/1.0/content/footer-6750a9
11 <script type="text/javascript" async="" charset="utf-8" id="utag_ancestry.main_1424" src="//tags.tiqcdn.com/utag/ancestry/main/prod/ut
12
13 <body style="position: relative" id="site-wrapper">
14 <div id="BannerRegion" style="position: relative; z-index: 9990;"></div>
15 <div id="HeaderRegion">
16 <nav class="nav navLoggedIn navStandard" id="nav" aria-label="Main Menu" data-nav-version="navVersion21" data-nav-type="ancestry-sta
17 <a href="#skipped-nav" class="hideVisually" id="navSkip">Skip Ancestry main menu</a>
18 <button type="button" class="link navMobile navLink icon iconMenu show480 calloutTriggerNoArrow id="navMobile" aria-controls="n
19 <a class="navLink navLogo" data-tracking-name="Logo" href="https://www.ancestry.com"><li><a class="navLink navMainLink navHome calloutTriggerNoArrow hide768 show480" data-tracking-n
21 <ul class="navSecondaryMenu"><li id="navSubscriptionOptions"><a class="navLink navEmphasizedButton hide480" href="https://www.an
22 <div class="navMobileMenu calloutDomContent" id="navMobileMenu"><a class="calloutMenuTitle navMobileHome" data-tracking-name="Ho
23
24
25 <div class="navCartMenu navSubMenu calloutDomContent navVersion21" id="navCartMenu">
26 <ul>
27 <li><a class="navCartCheckout" href="#" id="navCartCheckout" data-tracking-name="Check Out">Check Out</a></li>
28 <li><button class="link navCartEmpty" id="navCartEmpty" type="button" data-tracking-name="Empty Cart">Empty Cart</button>
29 </ul>
30 </div>
31 </nav>
32 <div id="skipped-nav"></div>
33
34 <script id="navData" type="application/json">{"langResources":{"Account":"Account","Hint_PersonHint":"Person Hint","Hint_PersonHintF
35 </div>
36 <app-root ng-version="9.1.12"><div class="discoveryui-matches-app sidebarWrapper sidebarPush"><!--><!--><div><router-outlet></ro
37 </button></tab><tab classes="icon iconMap" _ngghost-cyt-c60="" class="ng-star-inserted"><button _ngcontent-cyt-c60="" type="button" role=
38 </button></tab><!--></nav></div><!--></div></tabs-component><!--><div role="tabpanel" class="tabContent active"><router-outlet></r
39
40 <div id="FooterRegion">
41 <footer class="footer clearfix" id="footer" role="contentinfo">
42 <div class="footerLinks clearfix">

```

Close Family



1st – 2nd Cousin

Mother's Side

13% shared DNA | 904 cM ⓘ

 Unlinked tree

[View match](#)



1st – 2nd Cousin

Mother's Side

12% shared DNA | 821 cM ⓘ

 No family tree

[View match](#)



1st – 2nd Cousin

7% shared DNA | 497 cM ⓘ

 No family tree

[View match](#)



1st – 2nd Cousin

6% shared DNA | 445 cM ⓘ

 No family tree

[View match](#)



1st – 2nd Cousin

Mother's Side

6% shared DNA | 421 cM ⓘ

 No family tree

[View match](#)





1st – 2nd Cousin

Mother's Side

13% shared DNA

name,guid,rick,tree_status

"Rick Langston","D0D54F99-5310-44ED-A368-64D2C25397AD",9999,"NONE"

"my mother","15815692-A8FF-4151-95DE-DEC38B8D53E1",3456,"Unlinked tree"

"my brother1","D5F5B69F-CD60-4E10-A150-DDFB816EC228",2639,"63 People"

"my brother2","28ABEDD7-E7F9-48FF-8813-3EFB23DACE09",2380,"63 People"

"my 1st cousin1","30438C2E-BF01-4A87-932B-E08BD8C4CE1B",904,"Unlinked tree"

"my 1st cousin2","CFCC6AC8-C1BA-4BDC-9504-3799E22B8ABD",821,"No trees"

"my 1C1R","0468455C-04D3-4E84-9AFF-ADAA97216F1B",497,"No trees"

"my 2C","BAF8A512-0B23-49EE-BF68-CC5FF050455E",445,"No trees"

"another 1C1R","8F07F6F2-E6B1-493E-B0B0-972AC31D3877",421,"No trees"



1st – 2nd Cousin

Mother's Side

12% shared DNA | 821 cM ⓘ



1st – 2nd Cousin

7% shared DNA | 497 cM ⓘ



No family tree

...

[View match](#)

1st – 2nd Cousin

6% shared DNA | 445 cM ⓘ



No family tree

...

[View match](#)

1st – 2nd Cousin

Mother's Side

6% shared DNA | 421 cM ⓘ



No family tree

...

[View match](#)



1st - 2nd Cousin

Mother's Side

13% shared DNA

name,guid,rick,tree_status

"Rick Langston","D0D54F99-5310-44ED-A368-64D2C25397AD",9999,"NONE"

"my mother","15815692-A8FF-4151-95DE-DEC38B8D53E1",3456,"Unlinked tree"

"my brother1","D5F5B69F-CD60-4E10-A150-DDFB816EC228",2639,"63 People"

"my brother2","28ABEDD7-E7F9-48FF-8813-3EFB23DACE09",2380,"63 People"

"my 1st cousin1","30438C2E-BF01-4A87-932B-E08BD8C4CE1B",904,"Unlinked tree"

"my 1st cousin2","CFCC6AC8-C1BA-4BDC-9504-3799E22B8ABD",821,"No trees"

"my 1C1R","0468455C-04D3-4E84-9AFF-ADAA97216E1B",445,"No trees"

"my 2C","BAF8A512-0B23-49EE-BF68-CC5FF050455E",445,"No trees"

"another 1C1R","8F07F6F2-E6B1-493E-B0B0-972AC31D3877",421,"No trees"



1st - 2nd Cousin

Mother's Side

12% shared DNA | 821 cM ⓘ



1st - 2nd Cousin

7% shared DNA | 497 cM ⓘ

No family tree

View match



1st - 2nd Cousin

6% shared DNA | 445 cM ⓘ

No family tree

View match



1st - 2nd Cousin

Mother's Side

6% shared DNA | 421 cM ⓘ

No family tree

View match

Rick Langston's DNA Matches

List Map

Filter by:

Unviewed

Common ancestors

Messed

Notes

Trees

Shared DNA

Groups

Search Sort

Parent/Child



Open Link in New Tab
Open Link in New Window
Open Link in New Private Window

Bookmark This Link

Save Link As...

Save Link to Pocket

Copy Link Location

Copy

Search Google for "Rick Langston's..."

Send Link to Device

Print Selection

View Selection Source

Inspect Accessibility Properties

Inspect Element (Q)



Take a Screenshot



Video DownloadHelper

A | 3,456 cM



Unlinked tree

View match

Full Sibling



ed DNA | 2,639 cM



Linked tree | 63 People



Common ancestor

View match



ed DNA | 2,380 cM



Linked tree | 63 People



Common ancestor

View match

Close Family



Unlinked tree

View match

Rick Langston's DNA Matches

List Map

Filter by:

Unviewed

Common ancestors

Messed

Notes

Trees

Shared DNA

Groups

Search Sort

Parent/Child



- Open Link in New Tab
- Open Link in New Window
- Open Link in New Private Window

- Bookmark This Link
- Save Link As...
- Save Link to Pocket
- Copy Link Location
- Copy
- Search Google for "Rick Langston's..."

- Send Link to Device
- Print Selection
- View Selection Source

Full Sibling



A | 3,456 cM

Unlinked tree

View match

Linked tree | 63 People

Common ancestor

View match

Shared DNA | 2,639 cM

Matches/compare/D0D54F99-5310-44ED-A368-64D2C25397AD/with/15815692-A8FF-4151-95DE-DEC38B8D53E1

Close Family

Matches/compare/D0D54F99-5310-44ED-A368-64D2C25397AD/with/15815692-A8FF-4151-95DE-DEC38B8D53E1

Unlinked tree

View match

Rick Langston's DNA Matches

Filter by: Unviewed Common ancestors Messaged Notes Trees Shared DNA Groups Search Sort

Parent/Child

Unlinked tree View match

A | 3,456 cM

Open Link in New Tab
Open Link in New Window
Open Link in New Private Window
Bookmark This Link
Save Link As...

name, guid, rick, tree status
"Rick Langston", "D0D54F99-5310-44ED-A368-64D2C25397AD", 9999, "NONE"
"my mother", "15815692-A8FF-4151-95DE-DEC38B8D53E1", 3456, "Unlinked tree"
"my brother1", "D5F5B69F-CD60-4E10-A150-DDFB816EC228", 2639, "63 People"

Send Link to Device
Print Selection
View Selection Source

ed DNA | 3,639 cM

hes/compare/D0D54F99-5310-44ED-A368-64D2C25397AD/with/15815692-A8FF-4151-95DE-DEC38B8D53E1

Close Family

ii-matches/compare/D0D54F99-5310-44ED-A368-64D2C25397AD/with/15815692-A8FF-4151-95DE-DEC38B8D53E1

Unlinked tree View match



You and 

Predicted relationship: [Parent/Child](#)
50% shared DNA | 3,456 cM across 30 segments

[+ Add to group](#)

[Add note](#)

Trees Ethnicity Shared Matches

Filter by: [Unviewed](#)

[Common ancestors](#)

[Messaged](#)

[Notes](#)

[Trees](#) ▾

[Groups](#) ▾

[Search](#) | [Sort](#) ▾

Full Sibling



Brother

Mother's Side

47% - 54% shared DNA | 2,639 cM ⓘ



[Linked tree](#) | 63 People



Common ancestor

[View match](#)

...



Brother

Mother's Side

42% - 49% shared DNA | 2,380 cM ⓘ



[Linked tree](#) | 63 People



Common ancestor

[View match](#)

...

Close Family



1st - 2nd Cousin

Mother's Side

13% shared DNA | 904 cM ⓘ



[Unlinked tree](#)

[View match](#)

...



1st – 2nd Cousin

Mother's Side

12% shared DNA | 821 cM ⓘ

No family tree

[View match](#)



1st – 2nd Cousin

Mother's Side

6% shared DNA | 421 cM ⓘ

No family tree

[View match](#)



1st – 2nd Cousin

Mother's Side

6% shared DNA | 409 cM ⓘ

Linked tree | 17 People

Common ancestor

[View match](#)



1st – 2nd Cousin

Mother's Side

5% shared DNA | 365 cM ⓘ

Linked tree | 87 People

Common ancestor

[View match](#)



2nd – 3rd Cousin

Mother's Side

5% shared DNA | 334 cM ⓘ

Linked tree | 11 People

[View match](#)





1st – 2nd Cousin

Mother's Side

12% shared DNA | 821 cM ⓘ



1st – 2nd Cousin

Mother's Side

6% shared DNA | 421 cM ⓘ



1st – 2nd Cousin

Mother's Side

6% shared DNA | 409 cM ⓘ



Linked tree | 17 People



Common ancestor

View match



1st – 2nd Cousin

Mother's Side

5% shared DNA | 365 cM ⓘ



Linked tree | 87 People



Common ancestor

View match



2nd – 3rd Cousin

Mother's Side

5% shared DNA | 334 cM ⓘ



Linked tree | 11 People

View match



name,guid,rick,tree_status

"Rick Langston","D0D54F99-5310-44ED-A368-64D2C25397AD",9999,"NONE"
"my mother","15815692-A8FF-4151-95DE-DEC38B8D53E1",3456,"Unlinked t
"my brother1","D5F5B69F-CD60-4E10-A150-DDFB816EC228",2639,"63 Peopl
"my brother2","28ABEDD7-E7F9-48FF-8813-3EFB23DACE09",2380,"63 Peopl
"my 1st cousin1","30438C2E-BF01-4A87-932B-E08BD8C4CE1B",904,"Unlink
"my 1st cousin2","CFCC6AC8-C1BA-4BDC-9504-3799E22B8ABD",821,"No tre
"my 1C1R","0468455C-04D3-4E84-9AFF-ADAA97216FDB",497,"No trees"
"my 2C","BAF8A512-0B23-49EE-BF68-CC5FF050455E",445,"No trees"
"another 1C1R","8F07F6F2-E6B1-493E-B0B0-972AC31D3877",421,"No trees"



1st – 2nd Cousin
Mother's Side
12% shared DNA | 821 cM ⓘ

name, guid, rick, tree_status
"Rick Langston", "D0D54F99-5310-44ED-A368-64D2C25397AD", 9999, "NONE"
"my mother", "15815692-A8FF-4151-95DE-DEC38B8D53E1", 3456, "Unlinked t
"my brother1", "D5F5B69F-CD60-4E10-A150-DDFB816EC228", 2639, "63 Peopl
"my brother2", "28ABEDD7-E7F9-48FF-8813-3EFB23DACE09", 2380, "63 Peopl
"my 1st cousin1", "30438C2E-BF01-4A87-932B-E08BD8C4CE1B", 904, "Unlink
"my 1st cousin2", "CFCC6AC8-C1BA-4BDC-9504-3799E22B8ABD", 821, "No tre
"my 1C1R", "0468455C-04D3-4E84-9AFF-ADAA97216F0B", 497, "No trees"
"my 2C", "BAF8A512-0B23-49FF-BF68-CC5FF050455F", 445, "No trees"
"another 1C1R", "8F07F6F2-E6B1-493E-B0B0-972AC31D3877", 421, "No trees"

821



1st – 2nd Cousin
Mother's Side
6% shared DNA | 421 cM ⓘ



1st – 2nd Cousin
Mother's Side
6% shared DNA | 409 cM ⓘ

Linked tree | 17 People
Common ancestor

View match



1st – 2nd Cousin
Mother's Side
5% shared DNA | 365 cM ⓘ

Linked tree | 87 People
Common ancestor

View match



2nd – 3rd Cousin
Mother's Side
5% shared DNA | 334 cM ⓘ

Linked tree | 11 People

View match

Part 1 Generated Output

- The program uses the CSV files
- A cross-reference of matches is shown
- And an “old school hash” - CRC - is introduced
- We now go to demo data

x	name	john	tree_status	crc	guid
1	John Smith	(SELF)	100 People	FA10E4D1	B31299DA-17FE-A7A7-4305-98E1E913C4DD
2	Mary Jones	1000	50 People	8487E349	C7D2A76C-B961-79A7-52E2-6B3E9531120A
3	Bill Johnson	900	No Trees	457C2C02	4F6DEF7C-E44D-778E-E35C-0AF9C6BD6A67
4	James Johnson	500	Locked	562EDDE4	6B43DF42-200D-0EA8-770A-D44B5532A566
5	Ann Peterson	400	Unlinked Tree	7BFC95BF	EB09F9EB-81AA-36B6-B613-F66FDF74EBB0
6	Randy Jones	300	20 People	BE1B9664	991FC5CD-E8A3-3AFC-3792-8E786D72696F
7	Angela Smith	300	2 People	6784051D	71108249-7089-DB7E-56E2-8CA2704B503C
8	Kim Madison	250	No Trees	73723DB3	E3E1ABDB-8069-E08E-AA75-BF4E77B3D66B
9	Peter Wilson	200	123 People	196D8AF7	263E6EF0-8D19-1CB6-4352-5C9D0CE0421C

x	name	john	tree_status	crc	guid
1	John Smith	(SELF)	100 People	FA10E4D1	B31299DA-17FE-A7A7-4305-98E1E913C4DD
2	Mary Jones	1000	50 People	8487E349	C7D2A76C-B961-79A7-52E2-6B3E9531120A
3	Bill Johnson	900	No Trees	457C2C02	4F6DEF7C-E44D-778E-E35C-0AF9C6BD6A67
4	James Johnson	500	Locked	562EDDE4	6B43DF42-200D-0EA8-770A-D44B5532A566
5	Ann Peterson	400	Unlinked Tree	7BFC95BF	EB09F9EB-81AA-36B6-B613-F66FDF74EBB0
6	Randy Jones	300	20 People	BE1B9664	991FC5CD-E8A3-3AFC-3792-8E786D72696F
7	Angela Smith	300	2 People	6784051D	71108249-7089-DB7E-56E2-8CA2704B503C
8	Kim Madison	250	No Trees	73723DB3	E3E1ABDB-8069-E08E-AA75-BF4E77B3D66B
9	Peter Wilson	200	123 People	196D8AF7	263E6EF0-8D19-1CB6-4352-5C9D0CE0421C

x	name	john	mary	tree_status	crc
1	John Smith	(SELF)	1000	100 People	FA10E4D1
2	Mary Jones	1000	(SELF)	50 People	8487E349
3	Bill Johnson	900	-	No Trees	457C2C02
4	James Johnson	500	-	Locked	562EDDE4
5	Ann Peterson	400	Y	Unlinked Tree	7BFC95BF
6	Randy Jones	300	-	20 People	BE1B9664
7	Angela Smith	300	Y	2 People	6784051D
8	Kim Madison	250	-	No Trees	73723DB3
9	Peter Wilson	200	Y	123 People	196D8AF7

x	name	john	bill	mary	tree_status	crc
1	John Smith	(SELF)	900	1000	100 People	FA10E4D1
2	Mary Jones	1000	.	(SELF)	50 People	8487E349
3	Bill Johnson	900	(SELF)	.	No Trees	457C2C02
4	James Johnson	500	Y	.	Locked	562EDDE4
5	Ann Peterson	400	.	Y	Unlinked Tree	7BFC95BF
6	Randy Jones	300	—	—	20 People	BE1B9664
7	Angela Smith	300	Y	<u>Y</u>	2 People	6784051D
8	Kim Madison	250	Y	.	No Trees	73723DB3
9	Peter Wilson	200	.	Y	123 People	196D8AF7

- The Leeds Method
- Repeat as needed

Part 2: Descendancy

- We start examining the trees of those matches
- We look for common ancestors among them
- We show descendancies for these ancestors
- This runs counter to many researchers' beliefs!

Josiah Smith 1850-1920 + Eliza Fairfield 1852-1932
Charles Smith 1883-1950 + Eleanor James 1887-1965
Randolph Smith 1925-1990 + Elizabeth Nelson 1930-1998
Angela Smith 1948-
Raeford Smith 1885-1946 + Wilma Adamson 1890-1962
Quincy Smith 1918-1995 + Mary Floyd 1923-2010
Mary Smith 1955- + Phillip Jones
Elizabeth Smith 1957- + Martin Walker
Frederick Walker 1990-
William Smith 1920-1990 + Lucille Avery 1922-2000
John Smith 1950-
Edward Smith 1952-

George Avery 1882-1940 + Ina McDonald 1886-1956
Diane Avery 1919-1997 + George Johnson 1923-1998
William "Bill" Johnson 1955- + Maria Simpson 1961-
James Johnson 1992-
Amelia Johnson 1996-
Kenneth Avery 1920-1930
Lucille Avery 1922-2000 + William Smith (see above)

Josiah Smith 1850-1920 + Eliza Fairfield 1852-1932|
Charles Smith 1883-1950 + Eleanor James 1887-1965
Randolph Smith 1925-1990 + Elizabeth Nelson 1930-1998
Angela Smith 1948- [6784051D]
Raeford Smith 1885-1946 + Wilma Adamson 1890-1962
Quincy Smith 1918-1995 + Mary Floyd 1923-2010
Mary Smith 1955- [8487E349] + Phillip Jones
Elizabeth Smith 1957- + Martin Walker
Frederick Walker 1990-
William Smith 1920-1990 + Lucille Avery 1922-2000
John Smith 1950- [FA10E4D1]
Edward Smith 1952-

George Avery 1882-1940 + Ina McDonald 1886-1956
Diane Avery 1919-1997 + George Johnson 1923-1998
William "Bill" Johnson 1955- [457C2C02] + Maria Simpson 1961-
James Johnson 1992- [562EDDE4]
Amelia Johnson 1996-
Kenneth Avery 1920-1930

Josiah Smith	2	Mary Jones	1000	.	(SELF)	50 People	8487E349
Charles	3	Bill Johnson	900	(SELF)	.	No Trees	457C2C02
Randolph	4	James Johnson	500	Y	.	Locked	562EDDE4

Angela Smith 1978- [8487E349]

Raeferd Smith 1885-1946 + Wilma Adamson 1890-1962

Quincy Smith 1918-1995 + Mary Floyd 1923-2010

Mary Smith 1955- [8487E349] + Phillip Jones

Elizabeth Smith 1957- + Martin Walker

Frederick Walker 1990-

William Smith 1920-1990 + Lucille Avery 1922-2000

John Smith 1950- [FA10E4D1]

Edward Smith 1952-

George Avery 1882-1940 + Ina McDonald 1886-1956

Diane Avery 1919-1997 + George Johnson 1923-1998

William "Bill" Johnson 1955- [457C2C02] + Maria Simpson 1961-

James Johnson 1992- [562EDDE4]

Amelia Johnson 1996-

Kenneth Avery 1920-1930

Josiah Smith	2	Mary Jones	1000	.	(SELF)	50 People	8487E349
Charles :	3	Bill Johnson	900	(SELF)	.	No Trees	457C2C02
Randolph	4	James Johnson	500	Y	.	Locked	562EDDE4

Angela Smith 1970- [8487E349]

Raeford Smith 1885-1946 + Wilma Adamson 1890-1962

Quincy Smith 1918-1995 + Mary Floyd 1923-2010

Mary Smith 1955- [8487E349] + Phillip Jones

Elizabeth Smith 1957- + Martin Walker

Frederick Walker 1990-

William Smith 1920-1990 + Lucille Avery 1922-2000

John Smith 1950- [FA10E4D1]

Edward Smith 1952-

George Avery 1882-1940 + Ina McDonald 1886-1956

Diane Avery 1919-1997 + George Johnson 1923-1998

William "Bill" Johnson 1955- [457C2C02] + Maria Simpson 1961-

James Johnson 1992- [562EDDE4]

Amelia Johnson 1996-

Kenneth Avery 1920-1930

Charles Smith 1883-1950 + Eleanor James 1887-1965
Randolph Smith 1925-1990 + Elizabeth Nelson 1930-1998
Angela Smith 1948- [john:300,bill:Y,mary:Y]
Raeford Smith 1885-1946 + Wilma Adamson 1890-1962
Quincy Smith 1918-1995 + Mary Floyd 1923-2010
Mary Smith 1955- [john:1000,bill:.,mary:(SELF)] + Phillip Jones
Elizabeth Smith 1957- + Martin Walker
Frederick Walker 1990-
William Smith 1920-1990 + Lucille Avery 1922-2000
John Smith 1950- [john:(SELF),bill:900,mary:1000]
Edward Smith 1952-

George Avery 1882-1940 + Ina McDonald 1886-1956
Diane Avery 1919-1997 + George Johnson 1923-1998
William "Bill" Johnson 1955- [john:900,bill:(SELF),mary:.] + Maria Simpson 1961-
James Johnson 1992- [john:500,bill:Y,mary:.]
Amelia Johnson 1996-
Kenneth Avery 1920-1930
Lucille Avery 1922-2000 + William Smith (see above)

x	name	john	bill	mary	tree_status	crc
*	1 John Smith	(SELF)	900	1000	100 People	FA10E4D1
*	2 Mary Jones	1000	.	(SELF)	50 People	8487E349
*	3 Bill Johnson	900	(SELF)	.	No Trees	457C2C02
*	4 James Johnson	500	Y	.	Locked	562EDDE4
	5 Ann Peterson	400	.	Y	Unlinked Tree	7BFC95BF
	6 Randy Jones	300			20 People	BE1B9664
*	7 Angela Smith	300	Y	Y	2 People	6784051D
	8 Kim Madison	250	Y	.	No Trees	73723DB3
	9 Peter Wilson	200	.	Y	123 People	196D8AF7

	Name	john	mary	bill	Tree status
<u>*</u>	John Smith	(SELF)	1000	900	100 People
<u>*</u>	Mary Jones	1000	(SELF)	.	50 People
<u>*</u>	Bill Johnson	900	.	(SELF)	No Trees
<u>*</u>	James Johnson	500	.	M	Locked
	Ann Peterson	400	F	.	Unlinked Tree
	Randy Jones	300	_	_	20 People
<u>*</u>	Angela Smith	300	F	M	2 People
	Kim Madison	250	.	M	No Trees
	Peter Wilson	200	F	.	123 People

SAS Code Highlights

- Gathering file metadata
- Creating CSV files from HTML
- Reading CSV files and merging them
- Computing CRCs
- Using a format to obtain data given a key

Gathering File Metadata

```
%macro file_metadata(dirname,metadata_dataset);  
%global opt_filename opt_lastmod opt_filesize;  
%let opt_filename = Filename;  
%let opt_lastmod   = Last Modified;  
%let opt_filesize  = File Size (bytes);  
options validvarname=any;
```

```
data &metadata_dataset;  
  length filename $256;  
  rc  = filename('dfileref', "&dirname.");  
  did = dopen('dfileref');  
  nmems = dnum(did);  
  do i=1 to nmems;  
    filename = dread(did,i);  
    output;  
  end;  
  rc = dclose(did);  
  rc = filename('dfileref');  
  keep filename;  
run;
```

```

data &metadata_dataset; set &metadata_dataset;
  fullqual = "&dirname./"||filename;
  length opt $32 optval $256;
  do while(1);
    fid = 0;
    filename_rc = filename('myfile',fullqual);
    if filename_rc ne 0 then leave;
    fid = fopen('myfile');
    put filename= fid=;
    if fid = 0 then leave;
    n_metadata = foptnum(fid);
    do i=1 to n_metadata;
      opt = foptname(fid,i);
      if opt = "&opt_filename." then continue;
      optval = finfo(fid,opt);
      output;
      end;
    leave;
  end;
  if fid ne 0 then rc = fclose(fid);
  if filename_rc = 0 then rc = filename('myfile');
  keep filename opt optval;
run;

```

```
proc transpose data=&metadata_dataset out=&metadata_dataset(drop=_NAME_);  
  by filename notsorted;  
  var optval;  
  id opt;  
run;  
%mend;
```

```
3          %file_metadata(testdir,myfiles);
```

```
...
```

```
11         data _null_; set myfiles;
```

```
12             put filename= "&opt_filesize."n= "&opt_lastmod."n=;
```

```
13         run;
```

```
filename=rick_mother_matches.txt 'File Size (bytes)'n=21 'Last Modified'n=03Feb2021:14:04:10
```

```
filename=rick_rick_matches.csv 'File Size (bytes)'n=22 'Last Modified'n=03Feb2021:14:04:33
```

```
filename=rick_rick_matches.txt 'File Size (bytes)'n=20 'Last Modified'n=03Feb2021:14:03:55
```

Creating CSV Files From HTML

```
%macro html2csv(filename,part1,part2,filesize);  
data _null_; infile "&dirname.&slash.&filename." recfm=f lrecl=&filesize. column=c missover;  
....  
    input @'batBeacon' @;  
    do while(c < &filesize.);  
        input @'with%2F' @;  
        if c < &filesize.  
            then input guid $char36. @;  
        end;
```

- Use RECFM=F and LRECL=discovered filesize
- Be sure to use COLUMN= and MISSOVER

Reading CSV files and merging them

```
%macro read_csv_and_merge(filename,first);  
proc import file="&filename." dbms=csv out=temp replace; run;  
proc sort; by guid name; run;  
%if &first %then %do;  
data all; set temp; run;  
%end; %else %do;  
data all; merge all temp(drop=tree_status); by guid name; run;  
%end;  
%mend read_csv_and_merge;
```

Reading CSV files and merging them

```
%macro read_csv_and_merge(filename,first);  
proc import file="%filename." dbms=csv out=temp replace; run;  
proc sort; by guid name; run;  
%if &first %then %do;  
data all; set temp; run;  
%end; %else %do;  
data all; merge all temp(drop=tree_status); by guid name; run;  
%end;  
%mend read_csv_and_merge;
```

```
name,guid,rick,tree_status  
| "Rick Langston", "D0D54F99-5310-44ED-A368-64D2C25397AD", 9999, "NONE"  
"my mother", "15815692-A8FF-4151-95DE-DEC38B8D53E1", 3456, "Unlinked tree"  
"my brother1", "D5F5B69F-CD60-4E10-A150-DDFB816EC228", 2639, "63 People"
```

Computing CRCs

```
crc = crcxx1(strip(guid)); /* Prior to 9.4m6, returns an integer */
```

```
crc = hashing('crc32',strip(guid)); /* As of 9.4m6, returns hex */
```

Using a format to obtain data given a key

```
data temp; set matches;
  fmtname='$cm';
  array cM _numeric_; /* john mary william ... */
  start=crc;
  length label $100;
  label=' ';
  do over cM;
    name=vname(cM);
    if int(cM) ne cM then cM=.y;
    label=cats(label, ',', vname(cM), ': ', cM);
  end;
  label=substr(label,2);
  keep fmtname start label;
run;

proc format cntlin=temp; run;
```

```

data temp; set matches;
  fmtname='$cm';
  array cM _numeric_; /* john bill mary ... */
  start=crc;
  length label $100;
  label=' ';
  do over cM;
    name=vname(cM);
    if int(cM) ne cM then cM=.y;
    label=cats(label,', ',vname(cM),': ',cM);
  end;
  label=substr(label,2);
  keep fmtname start label;
run;

```

```

William "Bill" Johnson 1955- [457C
James Johnson 1992- [562EDDE4]
Amelia Johnson 1996-

```

```

William "Bill" JOHNSON 1955- [john:500,bill:(SEI
James Johnson 1992- [john:500,bill:Y,mary:..]
Amelia Johnson 1996-

```

```
proc format cntlin=temp; run;
```

```

data temp; set matches;
  fmtname='$cm';
  array cm _numeric_; /* john mary william ... */
  start=crc;
  length label $100;
  label=' ';
  do over cm;
    name=vname(cm);
    if int(cm) ne cm then cm=.y;
    label=cats(label,', ',vname(cm),': ',cm);
  end;
  label=substr(label,2);
  keep fmtname start label;
run;

```

```
proc format cntlin=temp; run;
```

```

William "Bill" Johnson 1955- [457C
James Johnson 1992- [562EDDE4]
Amelia Johnson 1996-

```

```

William "Bill" JOHNSON 1955- [john:500,bill:(SEI
James Johnson 1992- [john:500,bill:Y,mary:..]
Amelia Johnson 1996-

```

```

crc = scan(_infile_,2,'[]');
_infile_ = scan(_infile_,1,'[]')||put(crc,$cm.)
||scan(_infile_,3,'[]');

```

Success Stories

- Some of my successes using this program

Problems with the non-SAS user community!

- No access to the SAS System without a license
- Well, there is University Edition...but...
- You need a VM, and a vApp, and lots of memory...
- And it dies in mid-2021!!!
- Maybe SAS On Demand for Academics
- Too much to ask of “non-techies”



Python as an implementation alternative

- It's been around since at least 2000
- Runs well on Windows and Unix and elsewhere
- ... Including the IBM mainframe!
- It's free!
- I used



Basic execution

- abc.py program / abc.sas program

```
x = 'Hello, '  
y = 'World!'  
print(x + y) |
```

```
|data _null_;  
  x = 'Hello, '  
  y = 'World!';  
  put x y;  
run;
```

Basic execution

- You can run in an IDE (e.g. Spyder)
- Similar to SAS Studio / EG / even DMS

- SAS processes DATA/PROC steps/macro code
- It compiles DATA steps / interprets PROC steps
- Python is specifically an interpreter

Loose typing

```
x = 'abc'           # it is a string
x = len(x)          # now it is an int
x = x > 1            # now it is a boolean
x = [x,x,x]          # now it is a list!
```

Strings

```
x = 'abc'  
x = x + x      # x is now abcabc  
y = x[0:2]     # y is now ab  
x = x.upper()  # upper method upcases
```

Lists and Dictionaries

```
x = [1,2,3]    # A collection of objects
print(x[0])    # print first element
x.append(4)    # add a new element
```

```
x = dict()
x['abc'] = 1
x['def'] = 2
x['ghi'] = 3
```

```
for each in x :
    print(each,x[each])
abc 1
def 2
ghi 3
```

No Macro Processor!

- Probably the biggest difference with SAS
- But there is import which can be very helpful

Python import vs. SAS

- pip install zlib
- Then your code has

```
import zlib  
...  
crc_value = zlib.crc32(line1,crc_value)
```

- Compare with %INCLUDE and Autocall...

Providing Python programs to users

- `pyinstaller --onefile mypgm.py`
- Creates `mypgm.exe` that is distributable
- It can be bloated though!
- But no Python environment needed by the user

More notes

- Python 2 vs. Python 3
- PANDAS

File Metadata

```
os.listdir('directory')  
os.stat('filename').st_mtime
```

Computing CRC

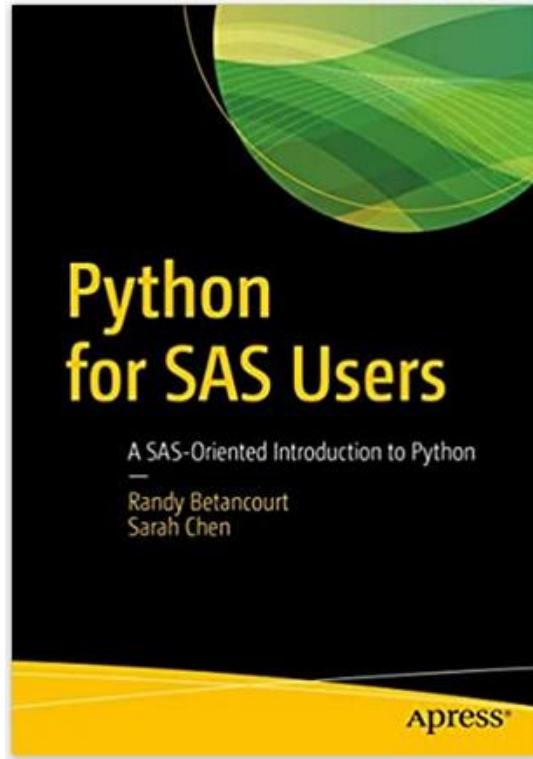
```
import zlib  
  
...  
crc_value = zlib.crc32(line1,crc_value)
```

Creating CSV from HTML

```
f = open('my.htm')
stream = f.read()
i = -1
while(True) :
    i = stream.find('/stuff/',i+1)
    if i == -1 : break
    ...
```

Using CRCs as keys

```
myinfo = dict()  
myinfo[crc] = 'John:1000;Mary:800;...'
```



#SASGF

SAS[®] GLOBAL FORUM 2021

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. * indicates USA registration. Other brand and product names are trademarks of their respective companies.

Summary

- AncestryDNA and its data
- My read_matches4 SAS program
- Reads data / organizes genealogical data
- Read_matches4 Part 2: researching the relatives
- Success in finding unknown relatives

Summary Continued

- Python as an alternative
- How it compares to SAS
- The program is available to the user community
- I will continue to use both SAS and Python

Thank you!

Contact Information
RickLangston1955@gmail.com