

Bioinformatics Institute (BII) Scientific Conference 2023

Creating value from
biomedical data

Sebastian Maurer -Stroh
ED BII
sebastianms@bii.a-star.edu.sg

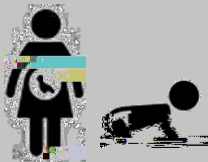
!"!#	\$%&'()*+ ,
!"!!	\$##"()*+ ,
!"!%	\$!%&'(-*+ ,

.+ / 0, 120(1230 /)402)(/ 54
 6*7(8*9*7) / (9* /)0:(; < (. ==

>20(/14?,0(+/@(6*7(+,,(*6(<*5A

B0)(5/(!"#\$%#&'\$(67&4(
*+,-#.+!\$(&.\$%\$%,/#%0#"

To address national health challenges & needs, A*STAR is increasingly directing efforts towards population health and well-being



!"#\$%&'()*+&
,+-. /&0234056789056:

Building on existing strengths in Early Life Research (e.g., S-PRESTO/GUSTO) Genomics (Precision Medicine) and Biodata capabilities (Biomed DAR), BMRC will expand our scope to address broader Population Health challenges for Singapore

A*STAR's BID A Simple Formula

!"#\$%&'%()*% '\$*#\$+,-.%

!"#\$%&'%\$"+%/0#\$"+1*#23%

4+%#*+%5)&21&6,%&6%788.

BII

Creating Value from Biomedical Data

!"\$%&'()*+,-./:;<=>?@A

Overall Goals in 5 years

A*STAR and BII as the Biomedical Data Analysis partner of choice for the whole ecosystem



BIODATA-highway

Why?

Biomedical data is very valuable (e.g. >\$110M value hosted in 2022)

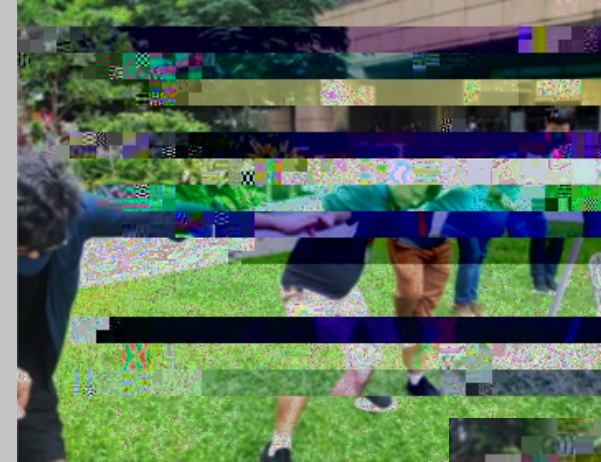
For who?

Singapore and the world

How?

Extend usable life span of high value biodata as secure datahub

Getting everyone on board



!"#\$%&&'

The only place with >100 Bioinformaticians within 4000km radius

BII

Critical mass needed to enable structured career path for Bio informaticians

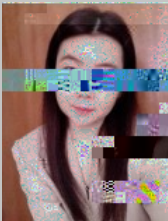
! 8<8C5*\$!&" ∅LL"*."3" /!,%(")6
! <%3,/!"!,B*!P&,*.& 0QCC=98=99QI86
! ?*\$ (4 /,3"/&AR("B 1%!&%+"4!,(/ ;*3)"!& 0S×E8I7:=CN:8CETG=SN8DETG=S6
! !"##\$%&#()""#*"+,-&'%-#.#. -%+".. /0. 1234""#&-55""#&\$\$. -%+".. 6 /0.7
! 8\$9 (+":5 ;\$-<. 2=22>=2?@



USS'V%(4'?%
0&"(%*!'F8:'I"&","!5;'
,/, '%(/"4!/,/%*(6



MXYYQ9T'Y\$+,!
0C,/,'95%"(5"
M,(34'6M8Y (:32 B)*(*)M%1Y



WNX'M,%'=;,(
0N&&/'F8:'9.,/%,3'
8++\$(*+%5&



?QS'T\$%'G%(4'7!,5"
0N&&/'F8:'988'1"33*B'
1*!'9.,/%,3'S+%5&6



Scott Mayfield



Linda Williams



V.J.G. Narayanan

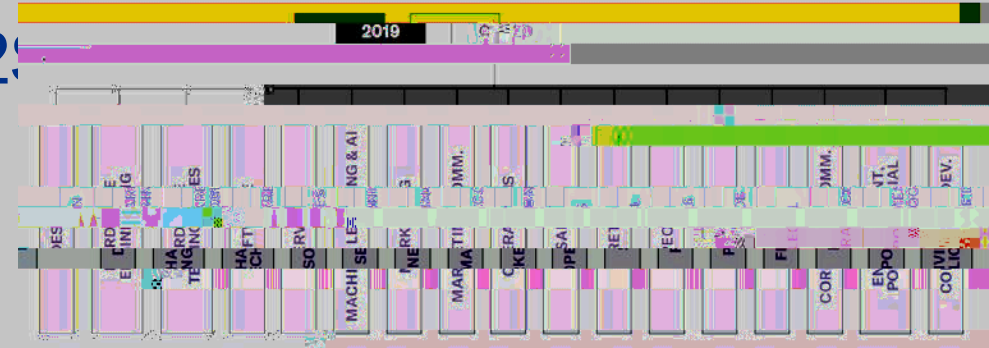
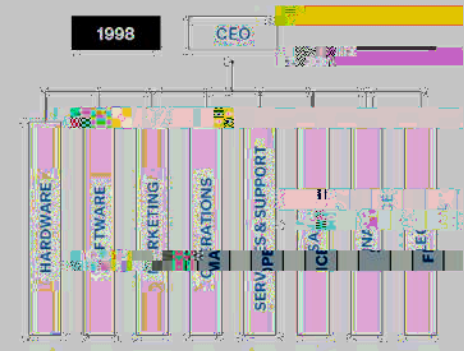


Rafael Di Tella

!"#\$%&'()*!#\$%\$&'()*&+&#&.--.\$/

! -. " = + ' 2 \$ # 5 " # \$) 3 7 \$ " : 2 \$ " " \$ & / (2 \$ 0 3 . / 2 # + # 3 : G \$
H 3 < \$ < 3 7 (> \$) 3 7 \$ " # # " 0 I \$ # 5 2 \$ 0 3 . / " ") 9 \$ & : 2
5 2 : 2 \$ " ") \$ < 2 " I ' 2 * * 2 * 9

! 4 5 " # \$ + 8 \$ & / (2 \$ < 3 7 (> \$ 6 7) \$) 3 7 : \$ 0 3 . / " ") A \$
< 5 " # \$ < 3 7 (> \$ # 5 2) \$ > 3 9



← W"&&

M*!" →



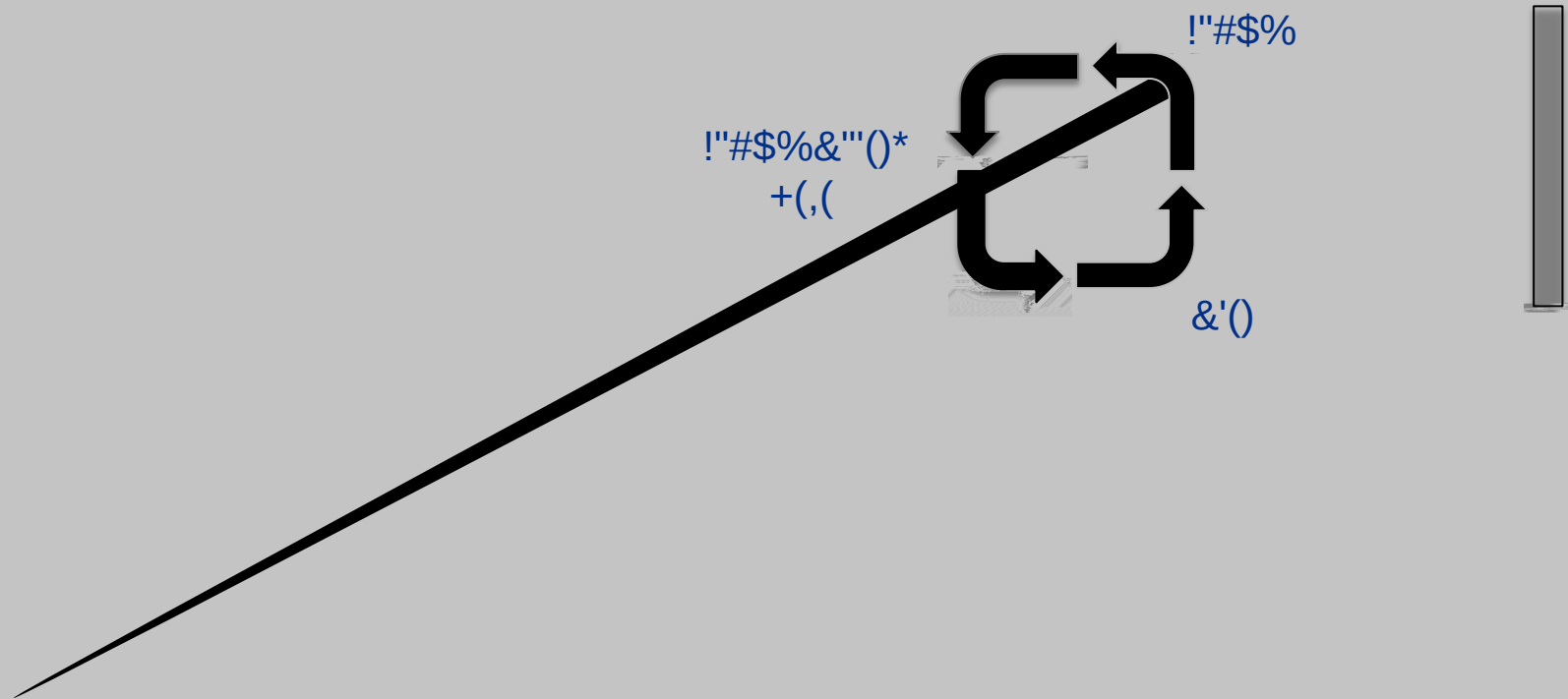
Simplicity is the ultimate sophistication

6%"44\$0,\$13545A0B/4\$,+4'0./"\$25*'2"1>",\$+,
.0074;EE???F&5)0)(/F15#E?"01.G=H<IJ+K3"L@:;>
16

How can we as BII increase the value created from data?



Strategy: Value Loop/Flywheel to grow Value Wedge



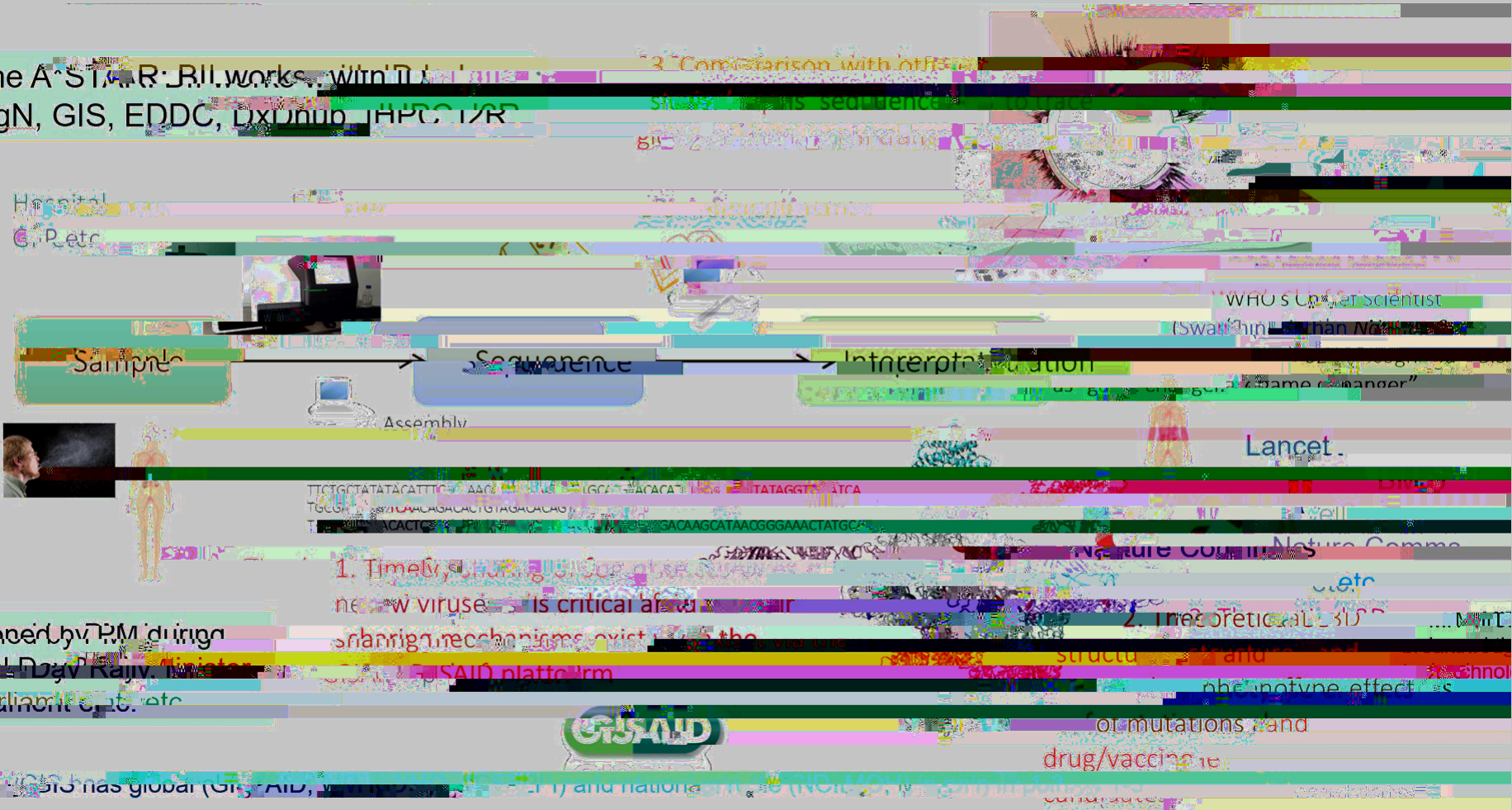
Adding Actionable Insights

BII

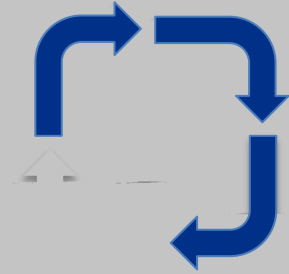
Creating Value from Biomedical Data

Example COVID-19:
data + analysis = impact

One A*STAR-BII works with
SIS, GIS, EDDC, DxHub, HPC, etc.



Value Loop/Flywheel to grow Value Wedge



Market -share is not equal to profitability and growth is a choice (raising prices can slow growth and increase profitability)

How to disrupt and survive disruption? Competitive Strategy

Blockbuster (physical DVD rental shops) vs Netflix (DVD-by-mail)

- ! Blockbuster : incumbent inertia, original business doing well, DVD-by-mail is ridiculous, you will never get to us
- ! Netflix (DVD -by -mail) :
 - ! indirect attack in area new to incumbent,
 - ! growing the market by reaching new customers (turning problem into virtue: latest releases less available, recommendation engine buffers load and suggests unexpected new diversity)



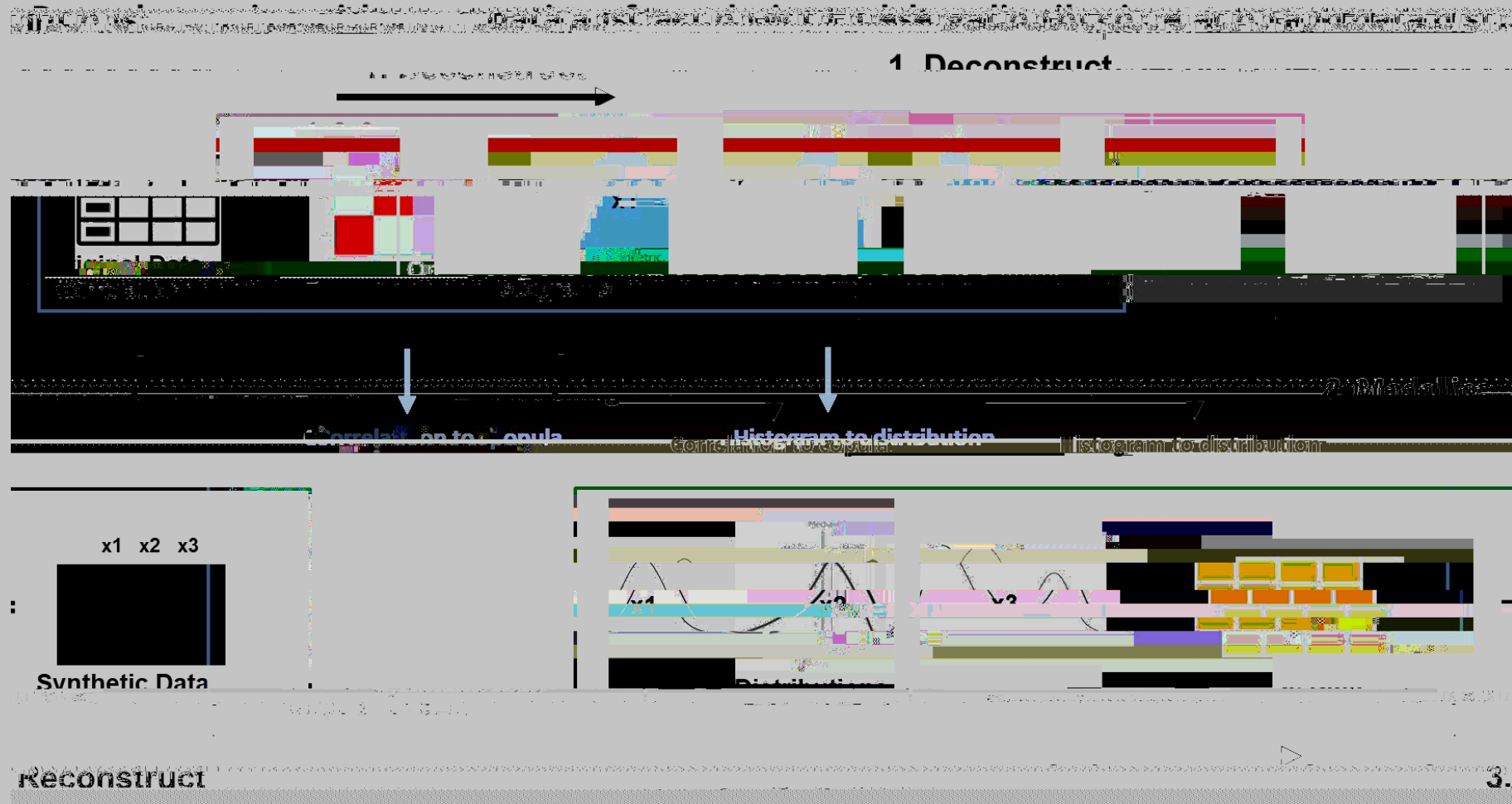
Patient re-identification is possible with access to few data elements

Regulatory, ethical and legal data protection hurdles

Extensive process is required for transferring and sharing of datasets for novel insights.

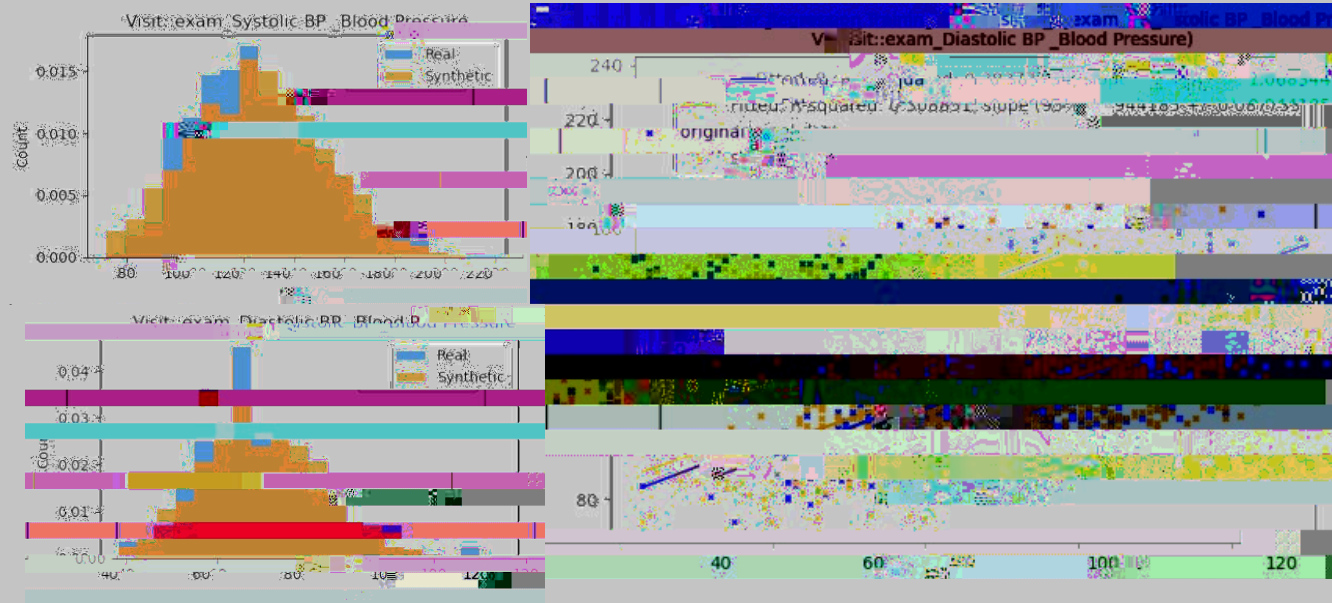
Heterogeneity of data increases difficulty of implementing high -

What if É we could remove sensitive aspects of the data but keep its biomedical value? Synthetic data



""-2')%+*)*:')A#2(*)E0*)'-2\$.)-%3):-K')%'B)*'1:%+\$+&.)'S,-%3#%&)*:')2'-\$#(7)+A)(. %*:'*#1)3-*-

Synthetic data is "real" (value for some purposes)



T+2)'S-7,\$'U)B')(:+B)+%) -)2'-\$)3-*-)(')*:-*)*:') ,2#K-1. H(-A')(.%*:'*#1)K'2(#+%)(*#\$)A#%3()
E-1V)*:')2'\$'K-%*)1\$#%#1-\$)1+22'\$-*#+% (=)

9:'2'A+2'U)(.%*:'*#1)3-*-)1-%)E')0('3)*+)A#%3)-%3)3'K'\$+,)7+3'\$()-E+0*)3#('-(')-%3)
*2' -*7'%*()A+2),-*#'%*()E0*)B#*:+0*)*:')(*2#%&()+A)%+*)E'#%&)-E\$)*+)(:-2')-%3)0(')B#3'\$.=
9:#():-()*:') ,+*'%*#-\$)%+*)+%\$.)*+)1+7,\$'7'%*()E0*)'K'%)*+)\$'-, HA2+&)1\$-((#1-\$)2'('-21:=

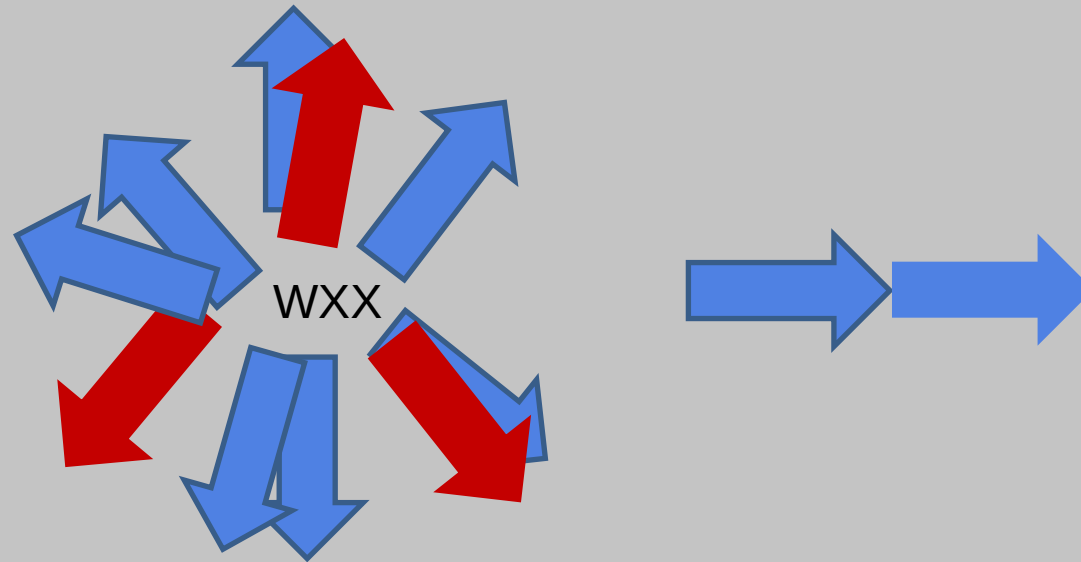


Computational complexity



!"#\$%&'()*#\$*%+\$,-.\$*)))\$-,\$\$/-01

2#3\$0%#\$%#&'\$()#&,-'\$
*+,-#.+!\$(&.\$%\$&% ,/#%0#"



10\$23&4,)5