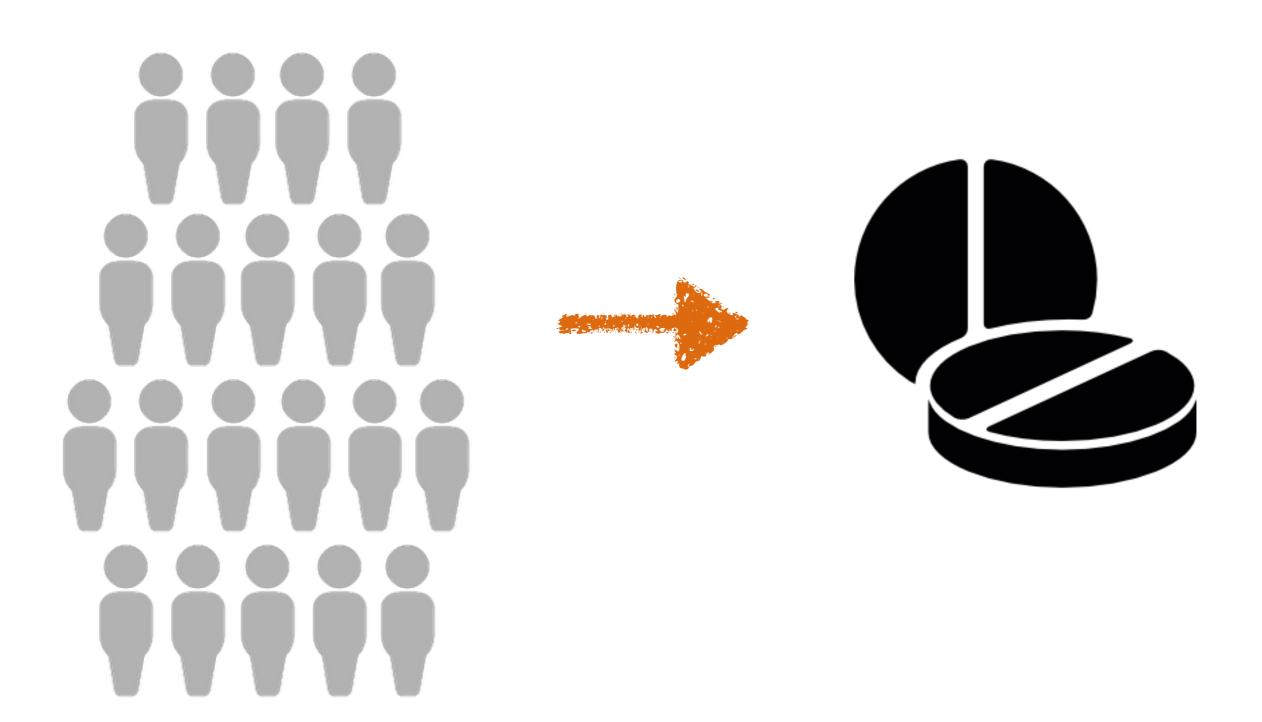
Understanding cancer behaviour with F#

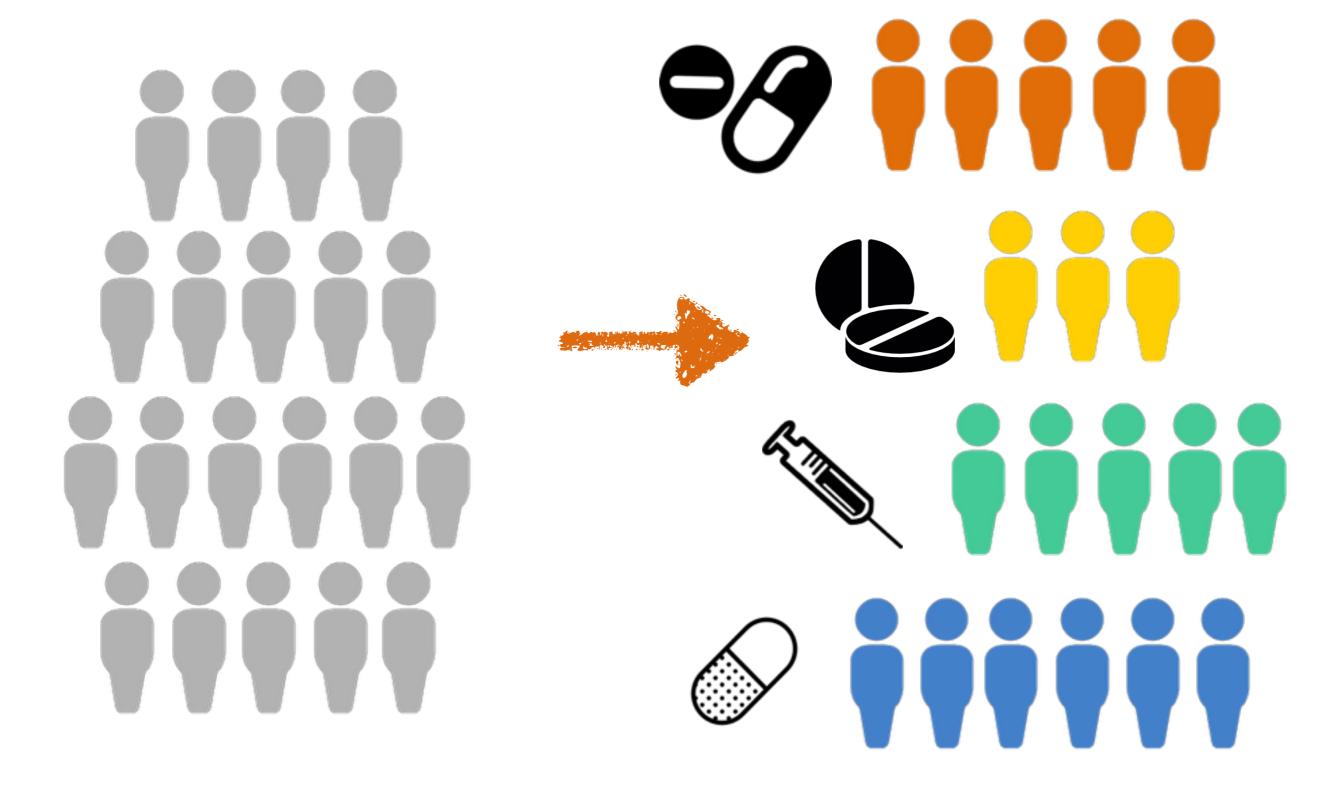
Evelina Gabasova

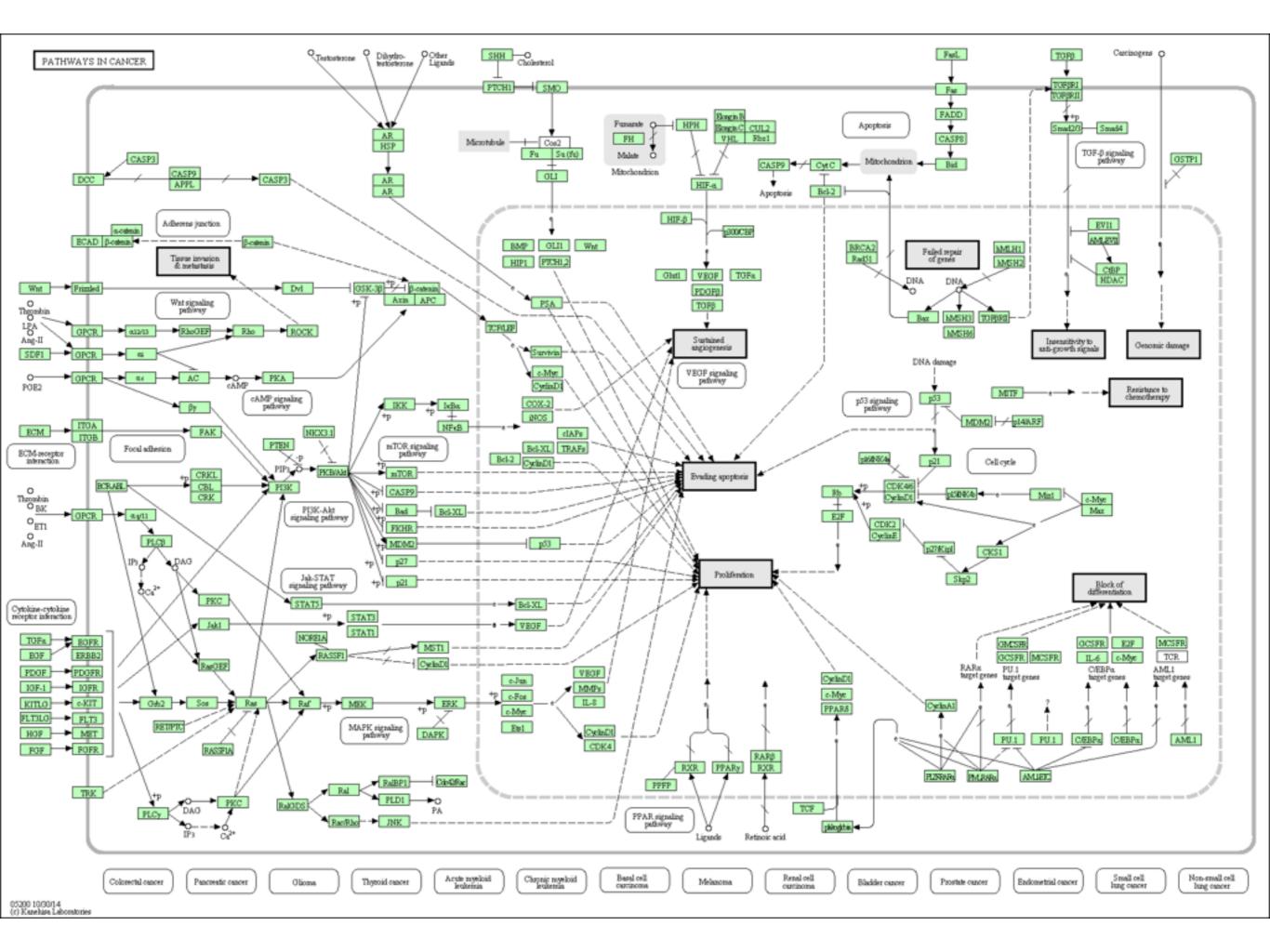
MRC Biostatistics Unit University of Cambridge

Precision medicine



Precision medicine





Integrative clustering

Breast cancer

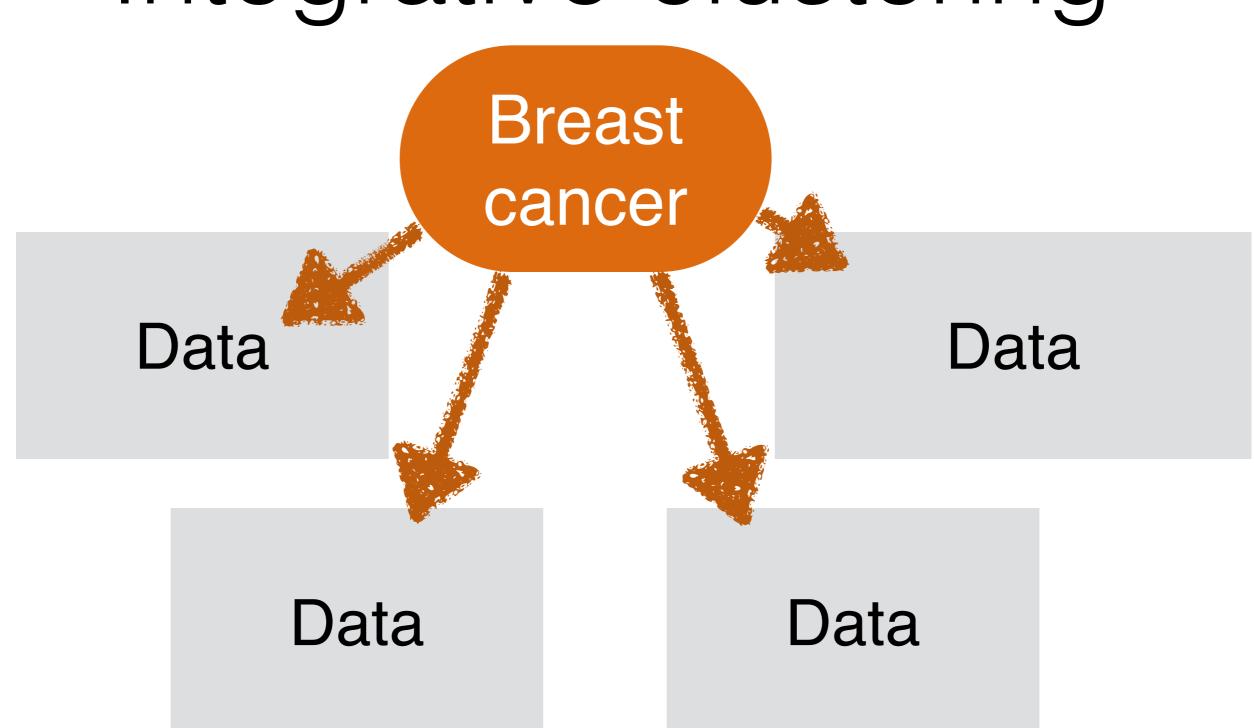
Gene expression

Reverse-phase protein array

microRNA expression

Methylation data

Integrative clustering



Scientific programming

Scripting languages

fast prototyping, easy to use

R, Python, Matlab

Why F#?

```
Lp = lapply(logL, exp)
pfor(m in 1:M) {
    for(i in 1:N) L[[m]][i,] = rmultinom(1,1,Lp[[m]][i,]) #Generate L from Lp
    if(w>1) L[[m]] = AlignClusters(C,L[[m]], type = 'mat') #Helps to align indices
    n[m,] = colSums(L[[m]])
    for (k in 1:K) { ###Update cluster parameters based on normal-gamma distribution
         if(d[m]==1&n[m,k]>1){
             S[[m]][,k] = sd(X[[m]][,L[[m]][,k]==1])^2
             PostMean = sum(X[[m]][,L[[m]][,k]==1])/(n[m,k]+1)
            B[[m]][,k] = b0[[m]]+0.5*(n[m,k]*S[[m]][,k]+n[m,k]*(mean(X[[m]][,L[[m]][,k]==1])-mu0
         if(d[m]>1&n[m,k]>1){
             PostMean = (mu0[[m]]+rowSums(X[[m]][,L[[m]][,k]==1]))/(n[m,k]+1)
             S[[m]][,k] = apply(X[[m]][,L[[m]][,k]==1],MARGIN=1,FUN='sd')^2
        B[[m]][,k] = b0[[m]]+0.5*(n[m,k]*S[[m]][,k]+n[m,k]*(rowMeans(X[[m]][,L[[m]][,k]==1])-mu0
         if(n[m,k]==1){
             PostMean = (mu0[[m]]+X[[m]][,L[[m]][,k]==1])/2
        B[[m]][,k] = b0[[m]]+0.5*(X[[m]][,L[[m]][,k]==1]-mu0[[m]])^2/2
         if(n[m,k]==0){
             PostMean = mu0[[m]]
            B[[m]][,k] = b0[[m]]
        Lambda = 1+n[m,k]
        A[[m]][,k] = a0[[m]]+n[m,k]/2
        Tau[[m]][,k] = rgamma(d[m],shape=A[[m]][,k],rate=B[[m]][,k])
        mu[[m]][,k] = rnorm(d[m], PostMean, sqrt(1/(Tau[[m]][,k]*Lambda)))
         Sigma[[m]][,k] = sgrt(1/Tau[[m]][,k])
```

Why F#?

```
Lp = lapply(logL, exp)
pfor(m in 1:M) {
    for(i in 1:N) L[[m]][i,] = rmultinom(1,1,Lp[[m]][i,]) #Generate L from Lp
    if(w>1) L[[m]] = AlignClusters(C,L[[m]], type = 'mat') #Helps to align indices
    n[m,] = colSums(L[[m]])
     for (k in 1:K) { ###Update cluster parameters based on normal-gamma distribution
         if(d[m]==1&n[m,k]>1){
             S[[m]][,k] = sd(X[[m]][,L[[m]][,k]==1])^2
             PostMean = sum(X[[m]][,L[[m]][,k]==1])/(n[m,k]+1)
             B[[m]][,k] = b0[[m]]+0.5*(n[m,k]*S[[m]][,k]+n[m,k]*(mean(X[[m]][,L[[m]][,k]==1])-mu0
         if(d[m]>1&n[m,k]>1){
             PostMean = (mu0[[m]]+rowSums(X[[m]][,L[[m]][,k]==1]))/(n[m,k]+1)
             S[[m]][,k] = apply(X[[m]][,k]==1], MARGIN=1, FUN='sd', ^2
         B[[m]][,k] = b0[[m]]+0.5*(n[k,l])*
         if(n[m,k]==1){
             PostMean = (mu0[[m]]+X[[m]][,L[[m]][,k]==1])/2
        B[[m]][,k] = b0[[m]]+0.5*(X[[m]][,L[[m]][,k]==0]-mu0[[m]])^2/2}
if(n[m,k]==0){

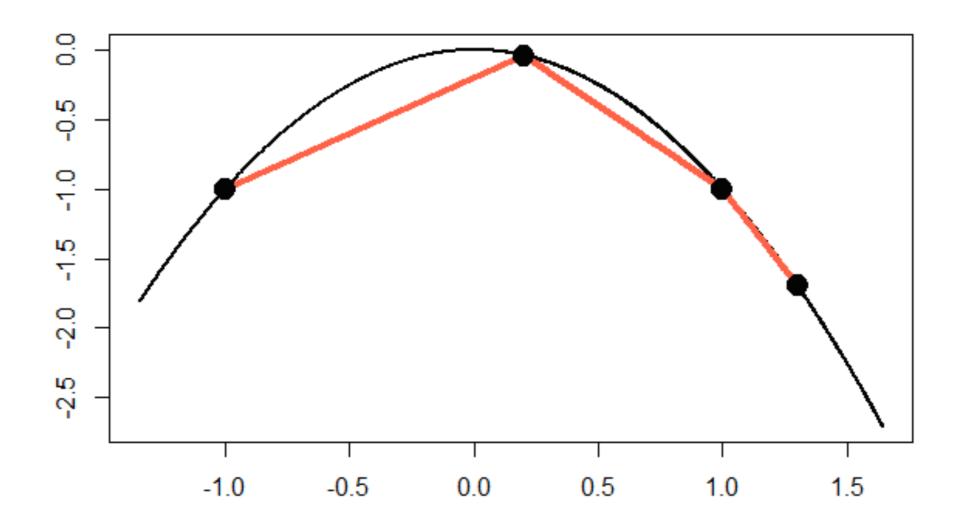
PostMean = mu0[[m]]
             PostMean = mu0[[m]]
             B[[m]][,k] = b0[[m]]
         Lambda = 1+n[m,k]
         A[[m]][,k] = a0[[m]]+n[m,k]/2
         Tau[[m]][,k] = rgamma(d[m],shape=A[[m]][,k],rate=B[[m]][,k])
         mu[[m]][,k] = rnorm(d[m], PostMean, sqrt(1/(Tau[[m]][,k]*Lambda)))
         Sigma[[m]][,k] = sgrt(1/Tau[[m]][,k])
```

Why F#?

```
25
           let newas =
26
               data.Contexts
27
               |> Array.map (fun context ->
                   let pis = state.ContextWeights.[context]
28
29
                   let rawPriorPis = Array.create pis.Length 1.0
                   let value = sampleDirichletConcentration (hyperprior.AlphasPrior.[context])
30
31
32
                    val sampleDirichletConcentration :
33
           { state w
                         gammaDist : Gamma ->
                                                                     as = newas \} 
                         rawPriorPis: seq<float> ->
34
       | FixedValue(
35
                                     : seq<float> ->
                         pis
36
           // do not
                         sampleType : SampleConcentrationParams
37
           state
                                    -> float
38
     Context-specif
                     Summary
     *********
                     Adaptive rejection sampling (derivative-free)
41
42 // Sample from conditional virience distribution with random watk Metropolis—Hastings
43 let sampleDirichlet_MetropolisHastings (currentValues: float[])
       priorConcentration loglikFunction =
44
45
      // 1. Add random walk proposal to current values
       let randomWalkProposal = Normal(0.0, 1.0, rnd)
46
       let proposal_unnorm =
47
48
           currentValues
           |> Array.map (fun x -> x + 0.1 * randomWalkProposal.Sample())
49
50
```

Case study

Rewrite an algorithm from Matlab into F# adaptive rejection sampling



Case study

```
% value of the lower bound
if x<min([lowerHull.left])</pre>
                                          0.5
  lhVal = -inf;
elseif x>max([lowerHull.right]);
  lhVal = -inf;
                                          <u>ا۔</u>
رن
else
                                          -2.0
  for li=1:length(lowerHull)
    left = lowerHull(li).left;
    right = lowerHull(li).right;
                                                  -1.0
                                                         -0.5
                                                               0.0
                                                                      0.5
                                                                             1.0
                                                                                   1.5
    if x>=left && x<=right
      lhVal = lowerHull(li).m*x + lowerHull(li).b;
      break;
    end
  end
          Matlab
end
```

Pattern matching

```
% value of the lower bound
                                    // value of the lower bound
                                    let lhVal =
if x<min([lowerHull.left])</pre>
  lhVal = -inf;
                                         match (lowerHull, x) with
                                           OutsideOnLeft -> -infinity
elseif x>max([lowerHull.right]);
                                           OutsideOnRight -> -infinity
  lhVal = -inf;
                                           InsideInterval lh -> lh.M * x + lh.B
else
  for li=1:length(lowerHull)
   left = lowerHull(li).left;
    right = lowerHull(li).right;
    if x>=left && x<=right
      lhVal = lowerHull(li).m*x + lowerHull(li).b;
      break;
   end
  end
         Matlab
end
```

Pattern matching

```
// value of the lower bound
% value of the lower bound
                                     let(lhVal) =
if x<min([lowerHull.left])</pre>
 lhVal = -inf;
                                         match (lowerHull, x) with
                                           OutsideOnLeft -> -infinity
elseif x>max([lowerHull.right]);
 lhVal = -inf;
                                           OutsideOnRight -> -infinity
                                           InsideInterval lh -> lh.M * x + lh.B
  for li=1:length(lowerHull)
    left = lowerHull(li).left;
    right = lowerHull(li).right;
    if x>=left && x<=right
     lhVal = lowerHull(li).m*x + lowerHull(li).b;
      break;
    end
  end
         Mattab
end
```

Legibility

$$X + \log \left(\sum_{i=1}^{N} \exp \left\{ x_i - X \right\} \right)$$
 where $X = \max \left\{ x_i; i = 1, \dots, N \right\}$

```
let logSumExp xs =
   let maxValue = Array.max xs
   xs
   |> Array.map (fun x -> exp(x - maxValue))
   |> Array.sum
   |> log
   |> (+) maxValue
```

Legibility

$$X + \log \left(\sum_{i=1}^{N} \exp \left\{ x_i - X \right\} \right)$$
where $X = \max \left\{ x_i; i = 1, \dots, N \right\}$

```
let logSumExp xs =
   let maxValue = Array.max xs
   xs
   |> Array.map (fun x -> exp(x - maxValue))
   |> Array.sum
   |> log
   |> (+) maxValue
```

Legibility

$$X + \log \left(\sum_{i=1}^{N} \exp \left\{ x_i - X \right\} \right)$$
 where $X = \max \left\{ x_i; i = 1, \dots, N \right\}$

```
let logSumExp xs =
   let maxValue = Array.max xs
   xs
   |> Array.map (fun x -> exp(x - maxValue))
   |> Array.sum
   |> log
   |> (+) maxValue
```

Problem

Breast cancer

Geneexpression

Revers

microRNA expression

Methylation data

Problem

$$\pi^{(d)} | \alpha_d, \beta, \pi \rangle$$

$$\propto p \left(\pi^{(d)} | \alpha_d \right) p \left(\pi | \pi^{(1)}, \dots, \pi^{(D)}, \beta \right)$$

$$= \left\{ \frac{\Gamma \left(K^{(d)} \alpha_d \right)}{\left(\Gamma(\alpha_d) \right)^{K^{(d)}}} \prod_{k=1}^{K^{(d)}} \left(\pi_k^{(d)} \right)^{\alpha_d - 1} \right\} \times$$

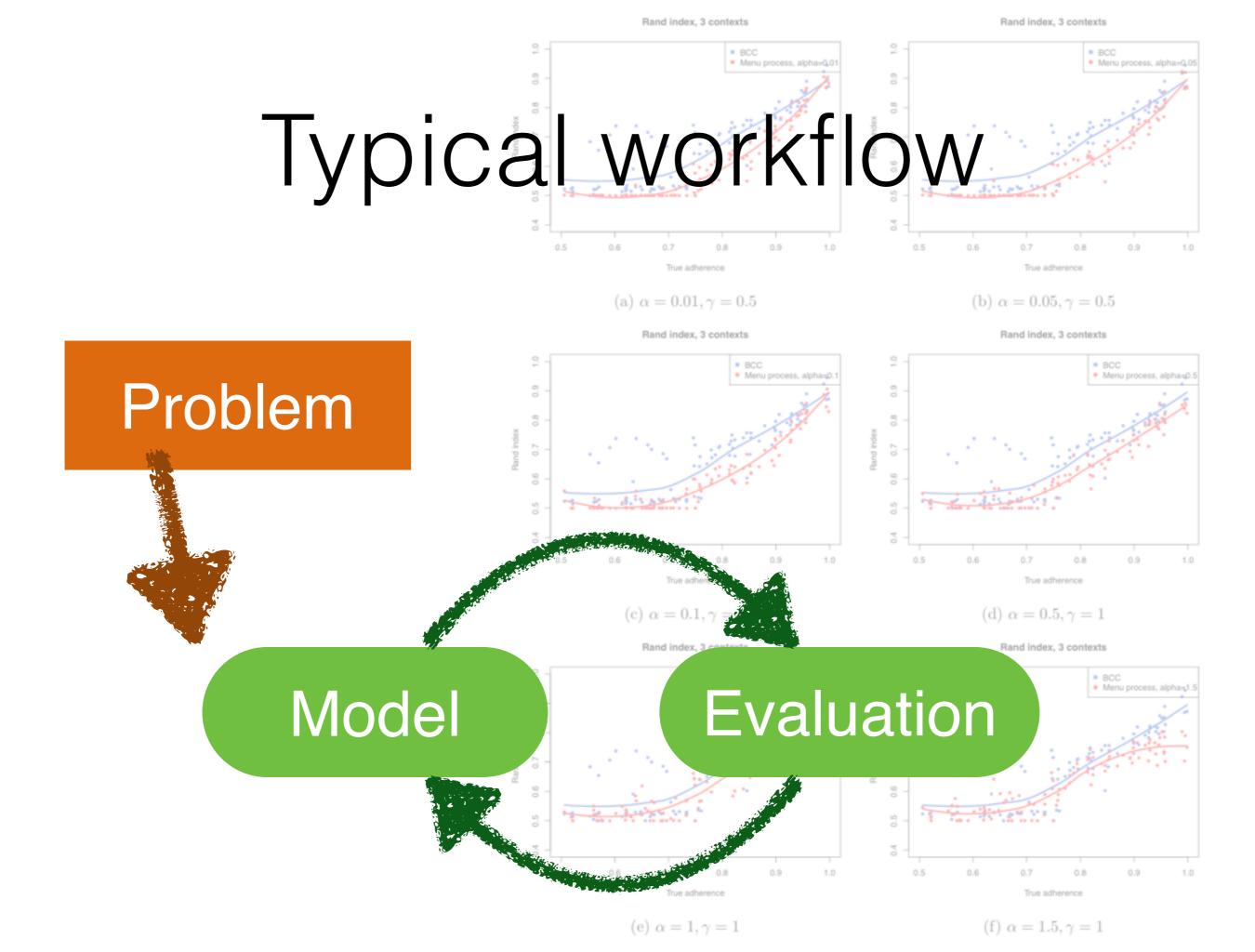
$$\times \left\{ \frac{\Gamma\left(\sum_{i_{1}=1}^{K^{(1)}} \cdots \sum_{i_{D}=1}^{K^{(D)}} \left[\beta \times \pi_{i_{1}}^{(1)} \times \cdots \times \pi_{i_{D}}^{(D)}\right]\right)}{\prod_{i_{1}=1}^{K^{(1)}} \cdots \prod_{i_{D}=1}^{K^{(D)}} \Gamma\left(\beta \times \pi_{i_{1}}^{(1)} \times \cdots \times \pi_{i_{D}}^{(D)}\right)} \prod_{i_{1}=1}^{K^{(1)}} \cdots \prod_{i_{D}=1}^{K^{(D)}} (\pi_{i_{1},...,i_{D}})^{\beta \pi_{i_{1}}^{(1)}...\pi_{i_{D}}^{(D)}-1} \right\}$$

 \mathbf{z}_{n}

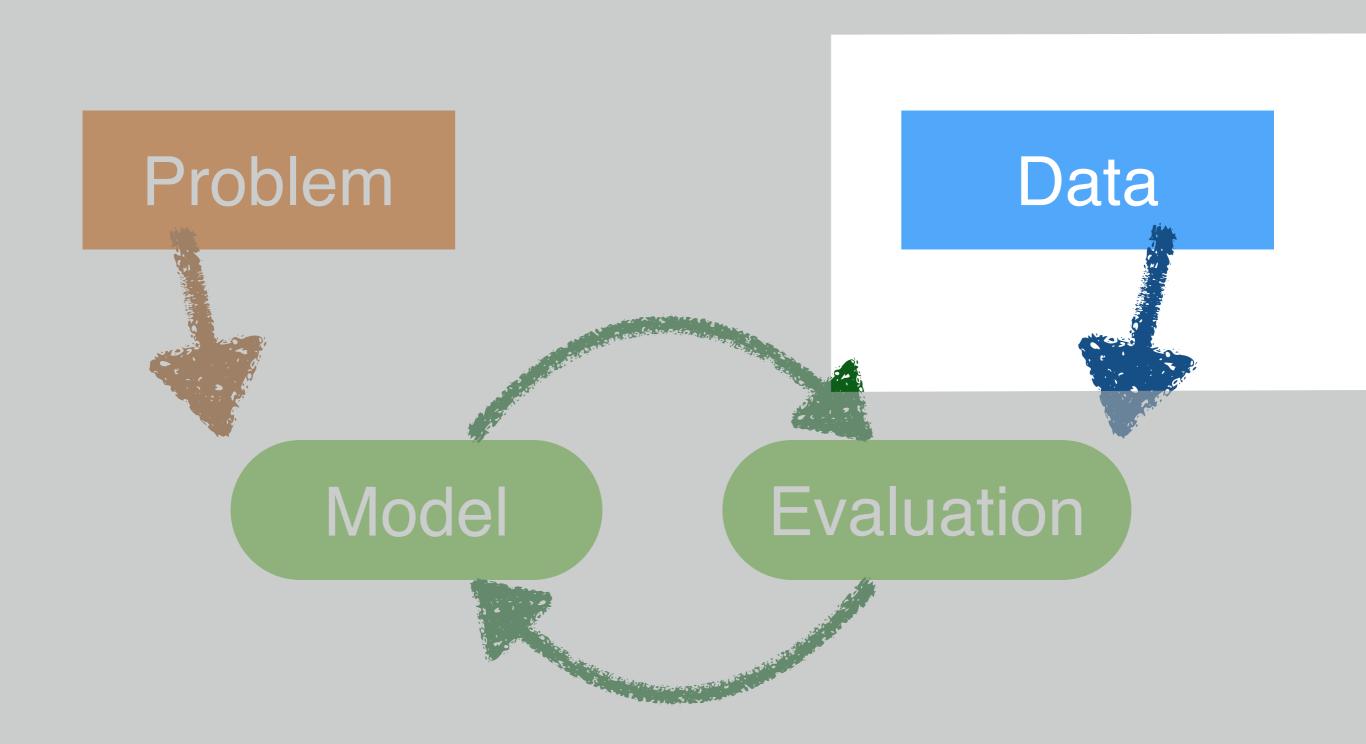
d = 1, ..., D

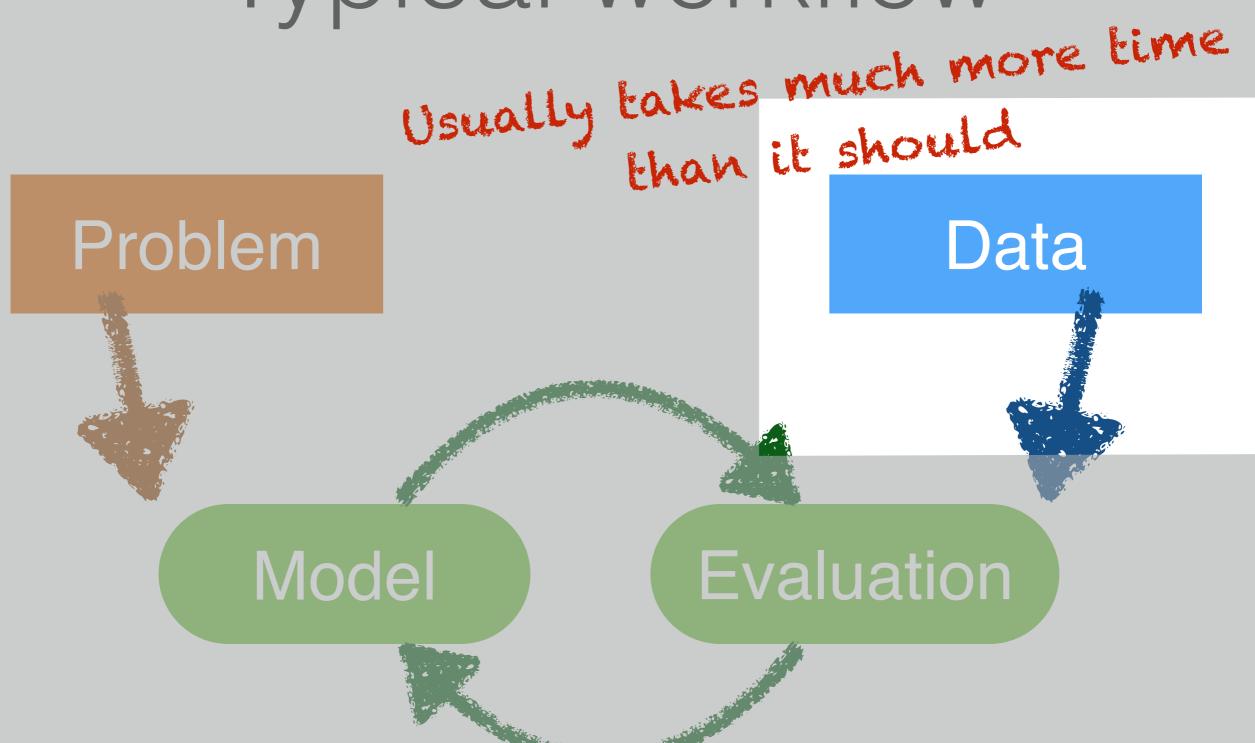
Model

$$\times \frac{\Gamma(\beta)}{\prod_{i_{1}=1}^{K^{(1)}} \cdots \prod_{i_{D}=1}^{K^{(D)}} \Gamma\left(\beta \times \pi_{i_{1}}^{(1)} \times \cdots \times \pi_{i_{D}}^{(D)}\right)} \times \\
\times \left\{ \prod_{k=1}^{K^{(d)}} \left(\pi_{k}^{(d)}\right)^{\alpha_{d}-1} \right\} \left\{ \prod_{i_{1}=1}^{K^{(1)}} \cdots \prod_{i_{D}=1}^{K^{(D)}} \left(\pi_{i_{1},...,i_{D}}\right)^{\beta \pi_{i_{1}}^{(1)}...\pi_{i_{D}}^{(D)}-1} \right\} \\
\propto \frac{\Gamma\left(K^{(d)}\alpha_{d}\right)}{\left(\Gamma(\alpha_{d})\right)^{K^{(d)}}} \left\{ \prod_{k=1}^{K^{(d)}} \left(\pi_{k}^{(d)}\right)^{\alpha_{d}-1} \right\} \left\{ \prod_{i_{1}=1}^{K^{(1)}} \cdots \prod_{i_{D}=1}^{K^{(D)}} \frac{1}{\Gamma\left(\beta \times \pi_{i_{1}}^{(1)} \times \cdots \times \pi_{i_{D}}^{(D)}\right)} \left(\pi_{i_{1},...,i_{D}}\right)^{\beta \pi_{i_{1}}^{(1)}...\pi_{i_{D}}^{(D)}-1} \right\}$$

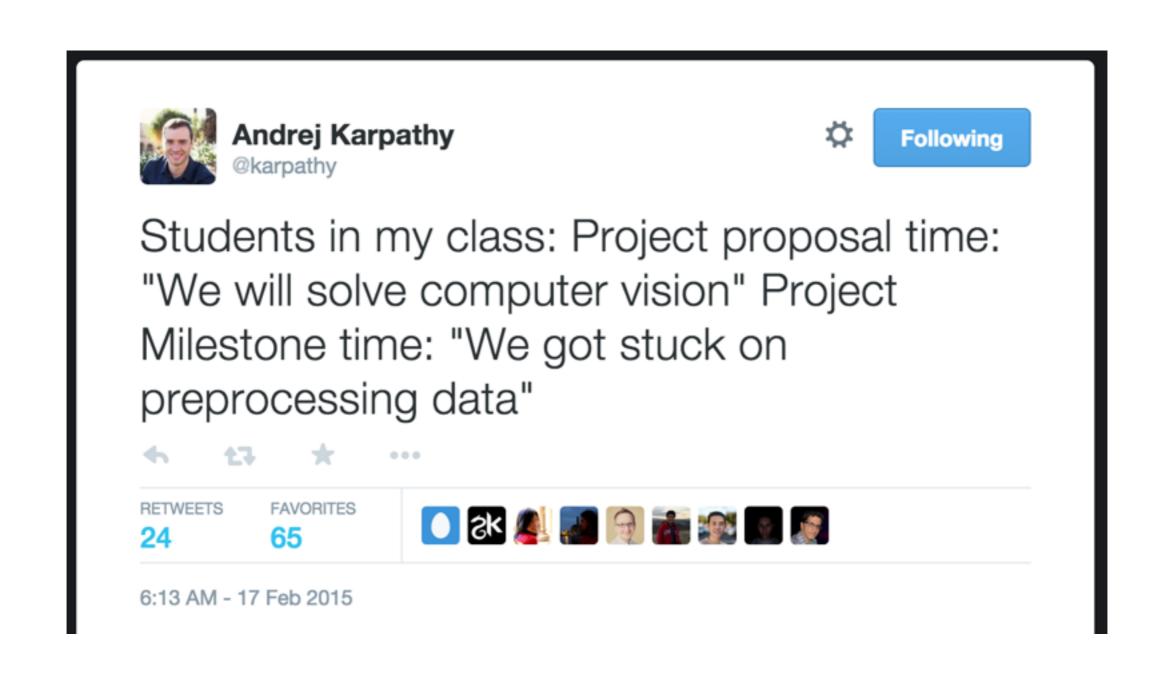


-A2-A0 FEMALE	36 Negative	Negative	Negative	T2	T_Other	N3	Positive	MO	Negative	Stage IIIC	No_Convers	i followup	DECEASED	547	<i>f</i>
-A1-A0! FEMALE	54 Negative	Negative	Negative	T2	T_Other	NO	Negative	MO	Negative	Stage IIA	Stage IIA	followup	DECEASED	594	k .
-A2-A0(FEMALE	40 Negative	Negative	Negative	T2	T_Other	NO	Negative	MO	Negative	Stage IIA	Stage IIA	followup	DECEASED	754	£
-AR-A1, FEMALE	50 Negative	Negative	Negative	T1	T1	N2	Positive	MO	Negative	Stage III	Stage IIIA	enrollment	DECEASED	[Not Availab	اد
-B6-A0\ FEMA				3	T_Other	N1	Positive	MO	Negative					653	1
-BH-A1 FEMA				1	T1	N1	Positive	MO	Negative					Availab	اد
-B6-A0I FEMA				1	T1	N1	Positive	MO	Negative					991	L.
-BH-A1 FEMA	Prob		\sim	2	T_Other	N1	Positive	MO	Negative			ata		1555	į.
-BH-A1 FEMA		ハモリ		2	T_Other	N1	Positive	MO	Negative					1692	1
-BH-A1 FEMA				1	T1	NO	Negative	M0	Negative					2547	
-BH-A0 FEMA	SAME.			2	T_Other	N1	Positive	MO	Negative			J	M.	72	2 NA
-BH-A0 FEMALE	rvegative	Negative	wegative	13	T_Other	N3	Positive	MO	Negative	Stage IIIC	NO_CONVERS	i ioliowup	· CIAIIAO	133	NA.
-BH-A0 FEMALE	Negative	Negative	Negative	T2	T_Other	NO	Negative	MO	Negative	Stage IIA	Stage IIA	followup	LIVING	170	NA C
-A7-A1: FEMALE	Negative	Positive	Negative	T2	T_Other	NO	Negative	MO	Negative	Stage IIA	Stage IIA	followup	LIVING	267	7 NA
-BH-A0 FEMALE	 Vegative 	Negative	Negative	T1	T1	NO	Negative	MO	Negative	Stage IA	Stage I	followurs	LIVING	292	NA S
-AO-A0 FEMALE	egative	Negative	Negative	T1	T1	Ma	P. Colonia Carlos	MO	Negative	Stage IA	Stage I	followue	LIVING	294	1 NA
-A7-A0(FEMALE	YES Y	Negative	Negative	T2	T_Other	Miss Con Contract	Netako	M0	Negative	Stage IIA	Stage IIA	6030	LIVING	309	NA E
-A7-A1: FEMALE		Negative	Negative	T2	T	N1	Positive	34	Negative	Stage IIB	No_Convers	,	NG	326	NA.
-A7-A0(FEMALE	i ne i je	Negative	Negative	T2	other	NO	Negative	M0	ogative	Stage IIA	Stage IIA		aMNG	373	NA.
-D8-A1 FEMALE	Taye	Negative	Negative	T3	T_Other	NO	Negative	MO	gative	Stage IIB	Stage IIB	fo	LIVING	425	5 NA
-D8-A1 FEMALE	5 Itive	Negative	Negative	T2	T_Other	NO	Negative	M0	gative	Stage IIA	Stage IIA	folk g	LIVING	431	l NA
-AQ-A0 FEMALE	45 Nopative	Negative	Negative	T2 1/2	T Other	NO	Negative	M0	ative	Stage IIA	Stage IIA	followap	LIVING	499	NA E
-BH-A0 FEMALE	67 Positive					NO	Negative					Vowup	LIVING	516	5 NA
-A2-A0' FEMALE	59 Negati					N1	Positive					vup	LIVING	533	NA.
-A2-A0' FEMALE	48 Nega				П	N1	Positiv			7		Ip.	LIVING	553	3 NA
-A2-A0' FEMALE	39 Posi				<i>A</i>	N2	Positiv			121		p	LIVING	565	5 NA
-A2-A0I FEMALE	60 Nega		Лoc			NO	Negati		valı			/p	LIVING	643	3 NA
-A2-A0 FEMALE	47 Negat					NO	Negativ					/up	LIVING	670	NA C
-AO-A0 FEMALE	61 Negative					NO	Negative					owup	LIVING	775	NA.
-A2-A0' FEMALE	67 Negative	Negative	Negative	12	II_Other	NO	Negative	MU	iw//nuve	Stage IIA	Stage IIA	followup	LIVING	964	NA.
-A2-A0I FEMALE	45 Negative	Negative	Negative	T2	ther	NO	Negative	MO	ative	Stage IIB	Stage IIA	followup	LIVING	1027	NA.
-BH-A0 FEMALE	53 Negative	Negative	Negative	T2	Other	N1	m tel	MO	Megative	Stage IIB	No_Convers	followup	LIVING	1203	NA.
-A2-A04 FEMALE	48 Negative	Negative	Negative	T1 6	L. A. B.	NO	Negative	M0	Negative	Stage IA	Stage I	followup	LIVING	1275	
-A2-A0! FEMALE	48 Negative	Negative	Negative	T1 3	P	NO	Negative	Mail	Negative	Stage IA	Stage I	followup	LIVING	1288	/ NA
-AO-A0 FEMALE	59 Negative	Negative	Negative	T2	T_Othe	E Comment	Positive		Negative	Stage IIIA	No_Convers		LIVING	1319	
-AO-A1 FEMALE	36 Negative	Negative	Negative	T2	T_Other	The Royal of the Park	Negative Negative Positive Negative Negative	MO	Negative	Stage IIA	Stage IIA	followup	LIVING	1471	
-BH-A0 FEMALE	44 Negative	Negative	Negative	T1	T1	NO	Negative	MO	Negative	Stage IA	Stage I	followup	LIVING	1572	

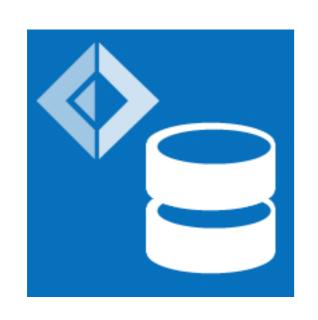




Dealing with real-world data



Type providers



F# Data - typed access to external data sources

Csv files

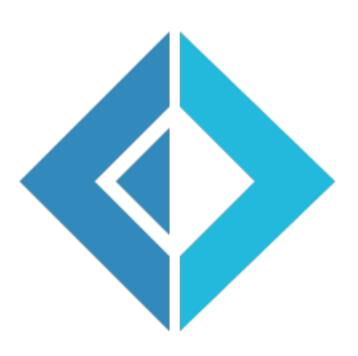
JSON, XML, HTML

RProvider



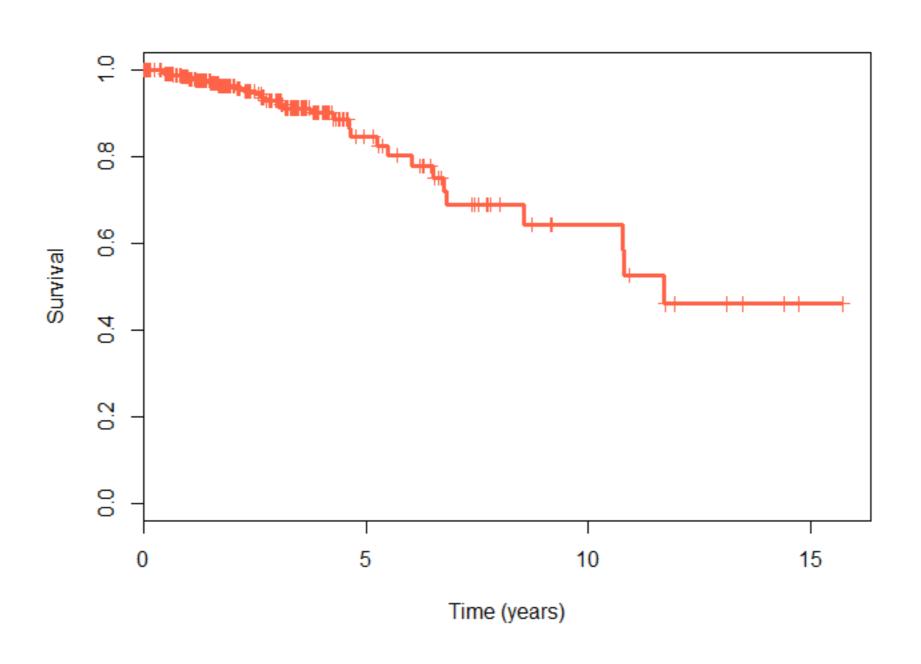




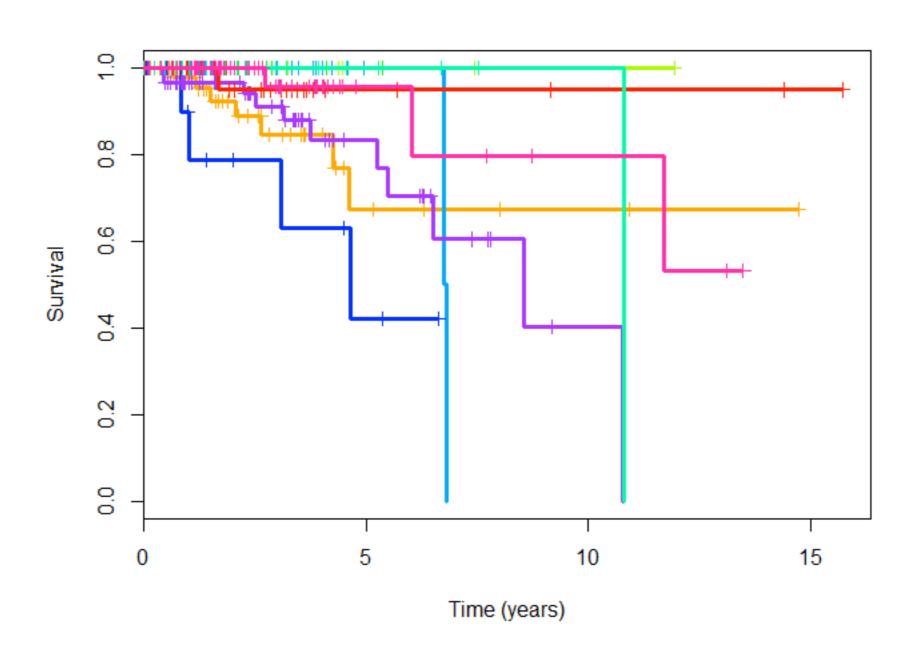




Results



Results



Thank you



@evelgab
evelinag.com
evelina@evelinag.com

F# eXchange 17 April in London