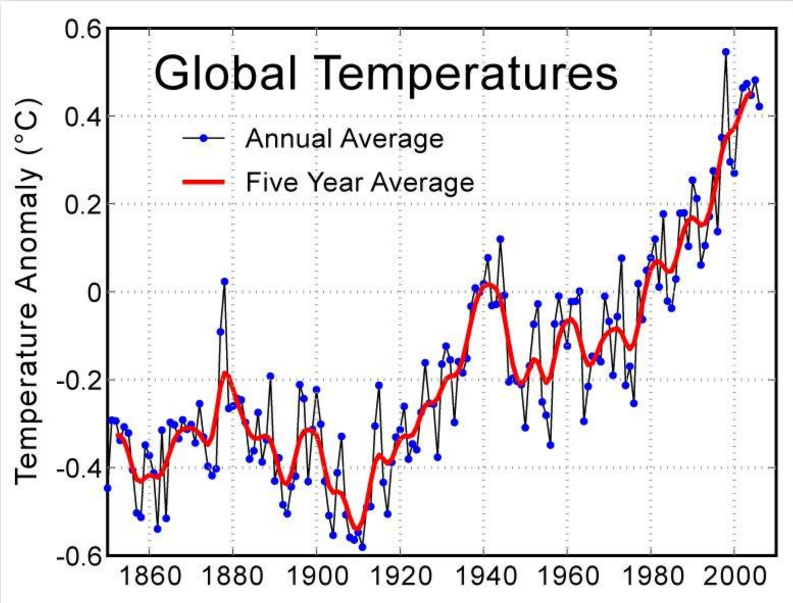
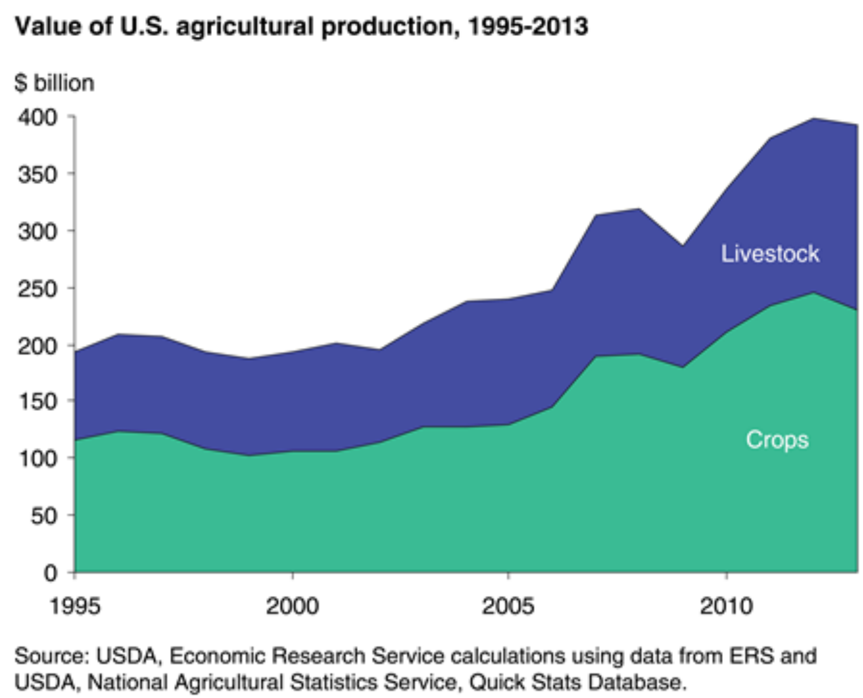
Accounting for GxE in Genomic Predictions for US Holstein dairy cattle

F Tiezzi, KL Parker-Gaddis, JS Clay and C Maltecca
North Carolina State University
USDA-ARS
DRMS

Background



- **'New' environmental conditions**



GxE in dairy cattle:

- Ravagnolo and Misztal, 2000
- Hayes et al., 2009
- Hammami et al., 2009
- Norberg et al., 2014
- Streit et al., 2013
- Windig et al., 2011
- Bryant et al., 2010
- Haile-Mariam et al., 2008
- Bohmanova et al., 2008
- Oseni et al., 2004
- Fikse et al., 2003
- Calus and Veerkamp, 2003
- Mulder and Bijma, 2005

Theor Appl Genet (2014) 127:595–607
DOI 10.1007/s00122-013-2243-1

ORIGINAL PAPER

A reaction norm model for genomic selection using high-dimensional genomic and environmental data

Diego Jarquín · José Crossa · Xavier Lacaze · Philippe Du Cheyron · Joëlle Daucourt · Josiane Lorgeou · François Piraux · Laurent Guerreiro · Paulino Pérez · Mario Calus · Juan Burgueño · Gustavo de los Campos

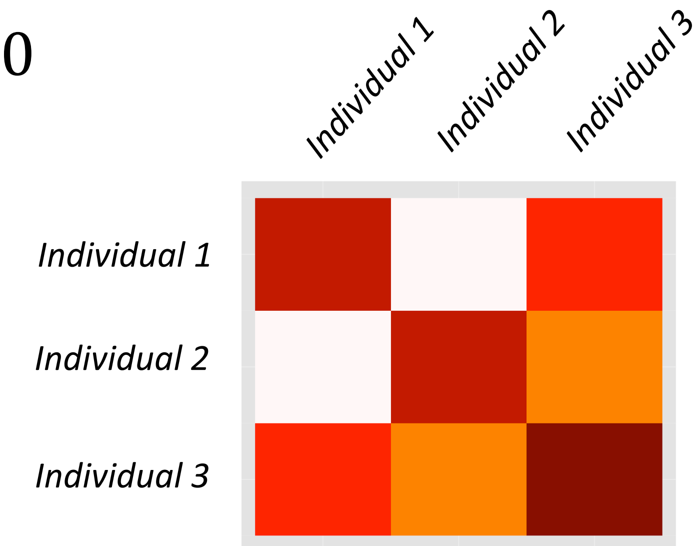
The Genomic Relationship Matrix

AA = 2
Aa = 1
aa = 0

Locus 1, Locus 2, Locus 3, Locus 4, Locus 5

$$\mathbf{M} = \begin{bmatrix} 1 & 2 & 2 & 0 & 1 \\ 0 & 0 & 2 & 1 & 1 \\ 0 & 2 & 2 & 1 & 0 \end{bmatrix}$$

Individual 1
Individual 2
Individual 3



$$\mathbf{G} = \frac{[\mathbf{M} - \mathbf{P}][\mathbf{M} - \mathbf{P}]'}{\sum 2p(1-p)}$$

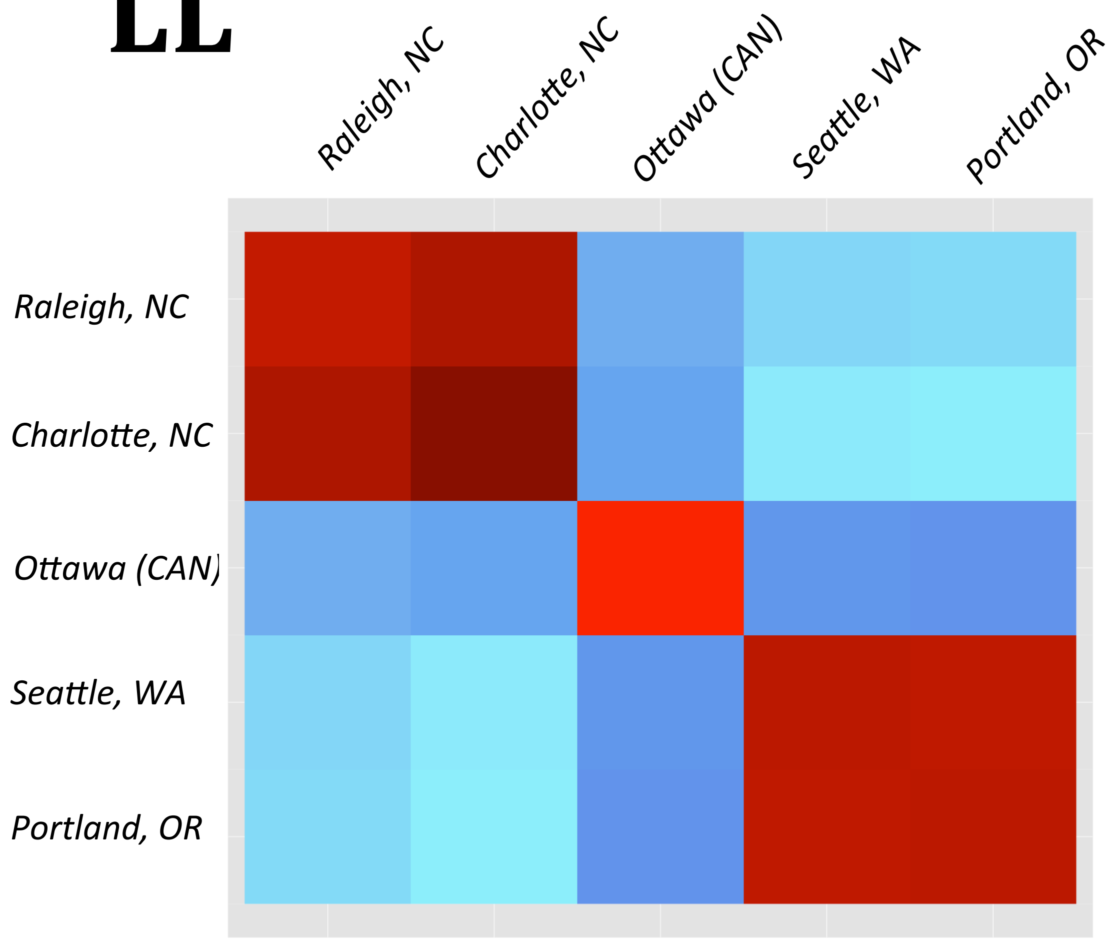
$$\mathrm{Cov}(\mathbf{g}) = \mathbf{G}\sigma_g^2$$

Nejati-Javaremi et al., 1997; VanRaden, 2008

The 'Geographical' Relationship Matrix

$$\mathbf{L} = \begin{bmatrix} 35.46 & 78.38 \\ 35.13 & 80.50 \\ 45.25 & 75.41 \\ 47.37 & 122.19 \\ 45.31 & 122.40 \end{bmatrix}$$

LL'

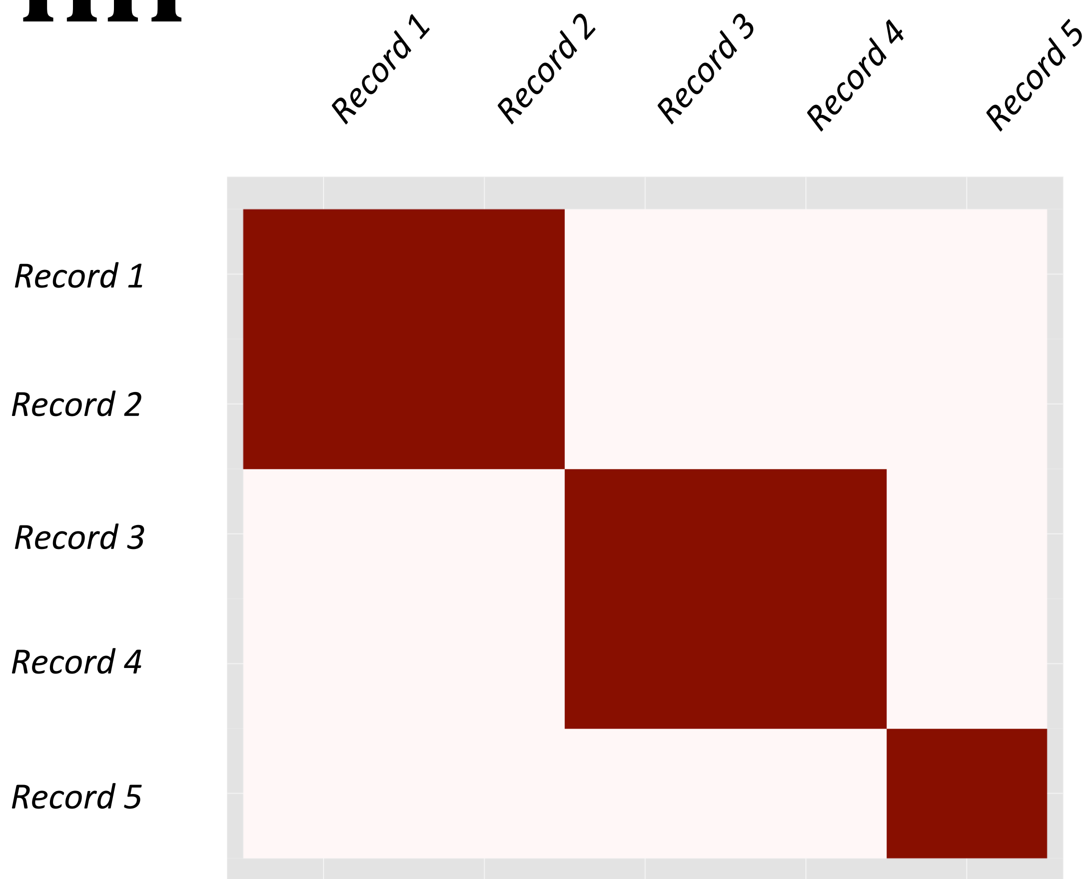


The 'Herd' Relationship Matrix

H=

	Herd 1	Herd 2	Herd 3
Record 1	1	0	0
Record 2	1	0	0
Record 3	0	1	0
Record 4	0	1	0
Record 5	0	0	1

HH'



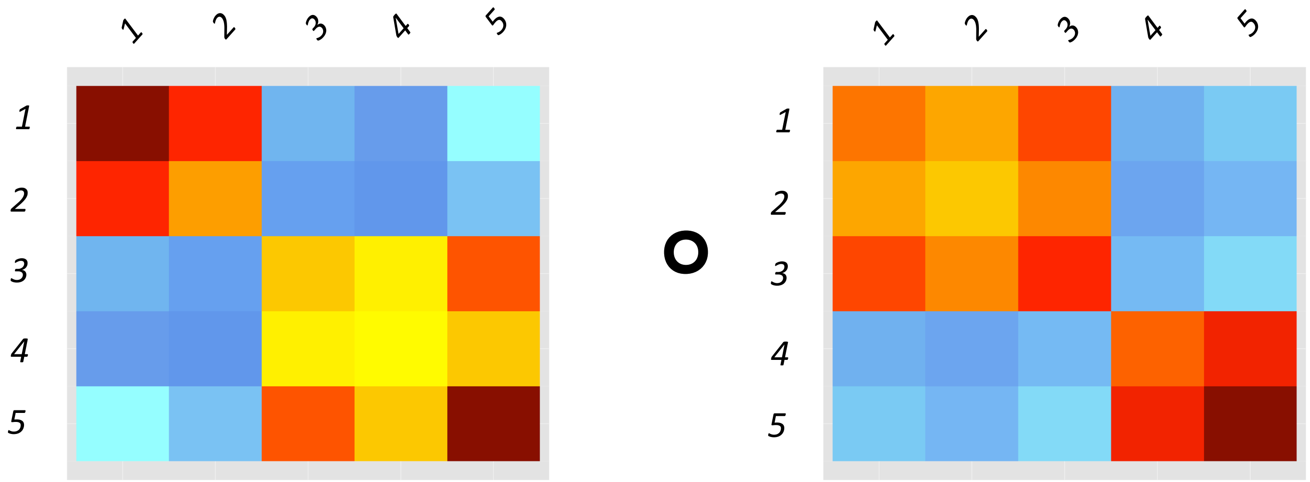
Hadamard product and the GxE

$$\begin{bmatrix} 2 & \cdots & 2 \\ \vdots & \ddots & \vdots \\ 2 & \cdots & 2 \end{bmatrix} \circ \begin{bmatrix} 3 & \cdots & 3 \\ \vdots & \ddots & \vdots \\ 3 & \cdots & 3 \end{bmatrix} =$$

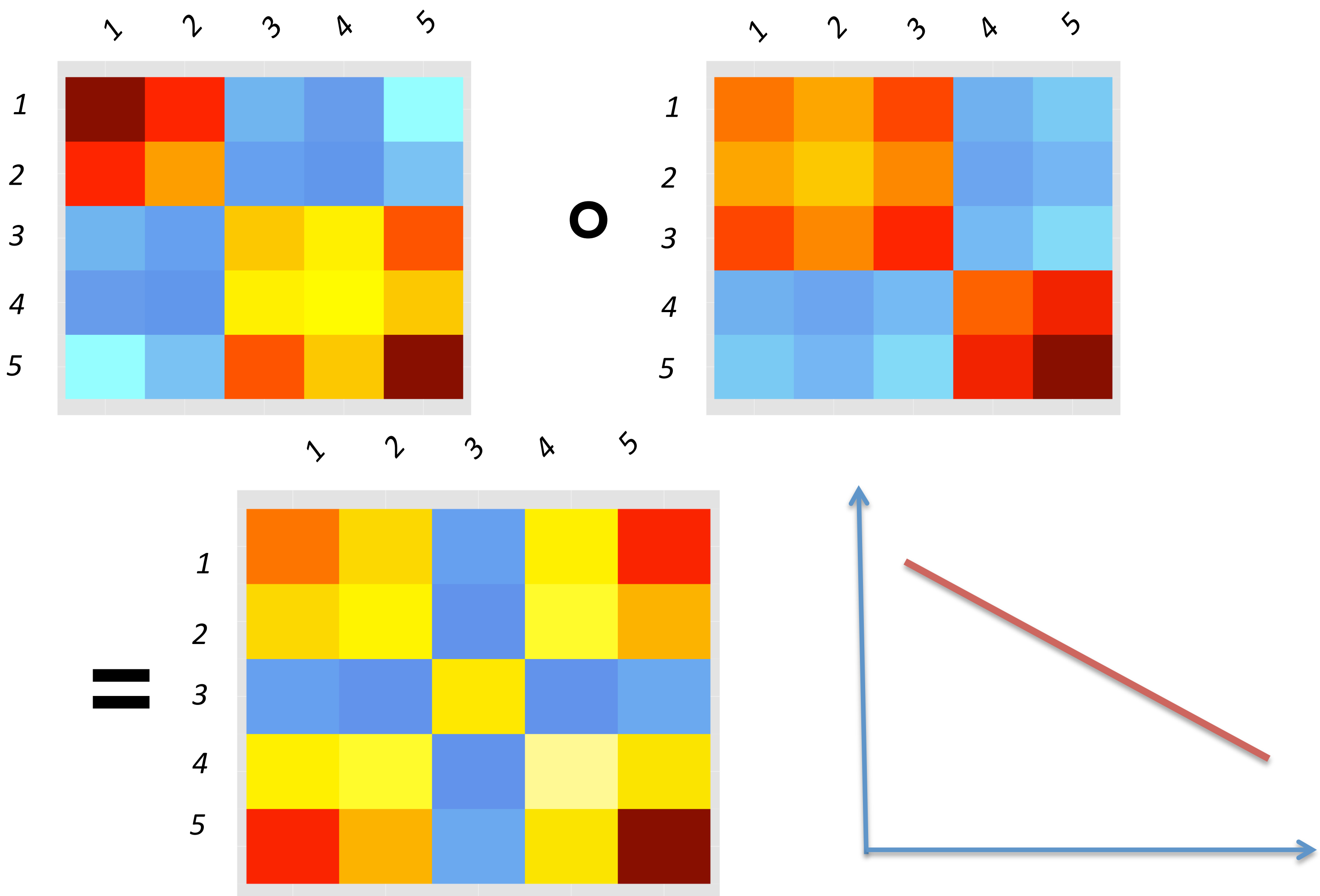
Hadamard product and the GxE

$$\begin{bmatrix} 2 & \cdots & 2 \\ \vdots & \ddots & \vdots \\ 2 & \cdots & 2 \end{bmatrix} \circ \begin{bmatrix} 3 & \cdots & 3 \\ \vdots & \ddots & \vdots \\ 3 & \cdots & 3 \end{bmatrix} = \begin{bmatrix} 6 & \cdots & 6 \\ \vdots & \ddots & \vdots \\ 6 & \cdots & 6 \end{bmatrix}$$

Hadamard product and the GxE



Hadamard product and the GxE



Materials and Methods

Data

- Test-day for ~2,000,000 cows
- Raised in 103 herds over 1,314 HYS classes
- Across 10 years
- 482 sires (54k genotypes)

- 11,747 within-HYS-DYDs for **Milk Yield**

Within-HYS-DYD

$$y_{ijklm} =$$
$$= \mu + parsolmf_i + hys_j + cowlact_{kl} + addgen_l + e_{ijklm}$$

$$DYD_{nj} = y_{ijklm} - parsolmf_i – cowlact_{kl}$$

$$= hys_j + addgen_l + e_{ijklm}$$

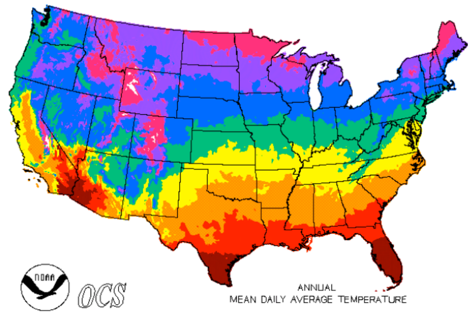
Data

- Test-day for ~2,000,000 cows
- Raised in 103 herds over 1,314 HYS classes
- Across 10 years
- 482 sires (54k genotypes)

- 11,747 within-HYS-DYDs for **Milk Yield**

- Latitude and Longitude
- NOAA Climate data
- Format 4 Management variables

Climate Data



- **Daily Measures**
  - Maximum Temperature, °C (19.5±9.0)
  - Minimum Temperature, °C (7.8±8.4)
  - Average Temperature, °C (13.8±8.7)
  - Relative Humidity, % (65.9±7.0)
  - Pressure, mmHg (744.9±21.9)
  - Wind Speed, km/h (9.9±3.0)

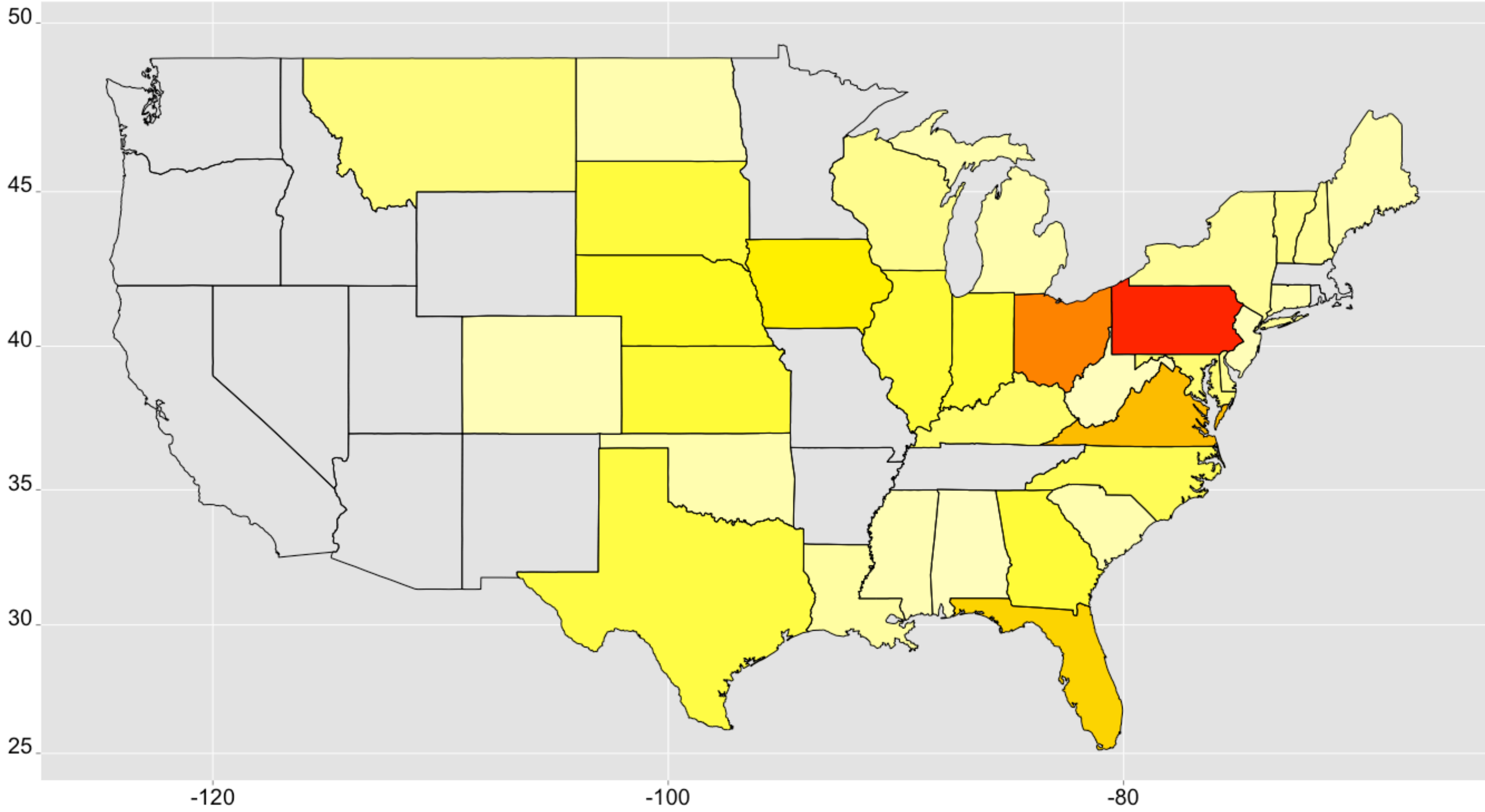
- **Monthly Sum**
  - Rainfall, mm (251.0±103.2)

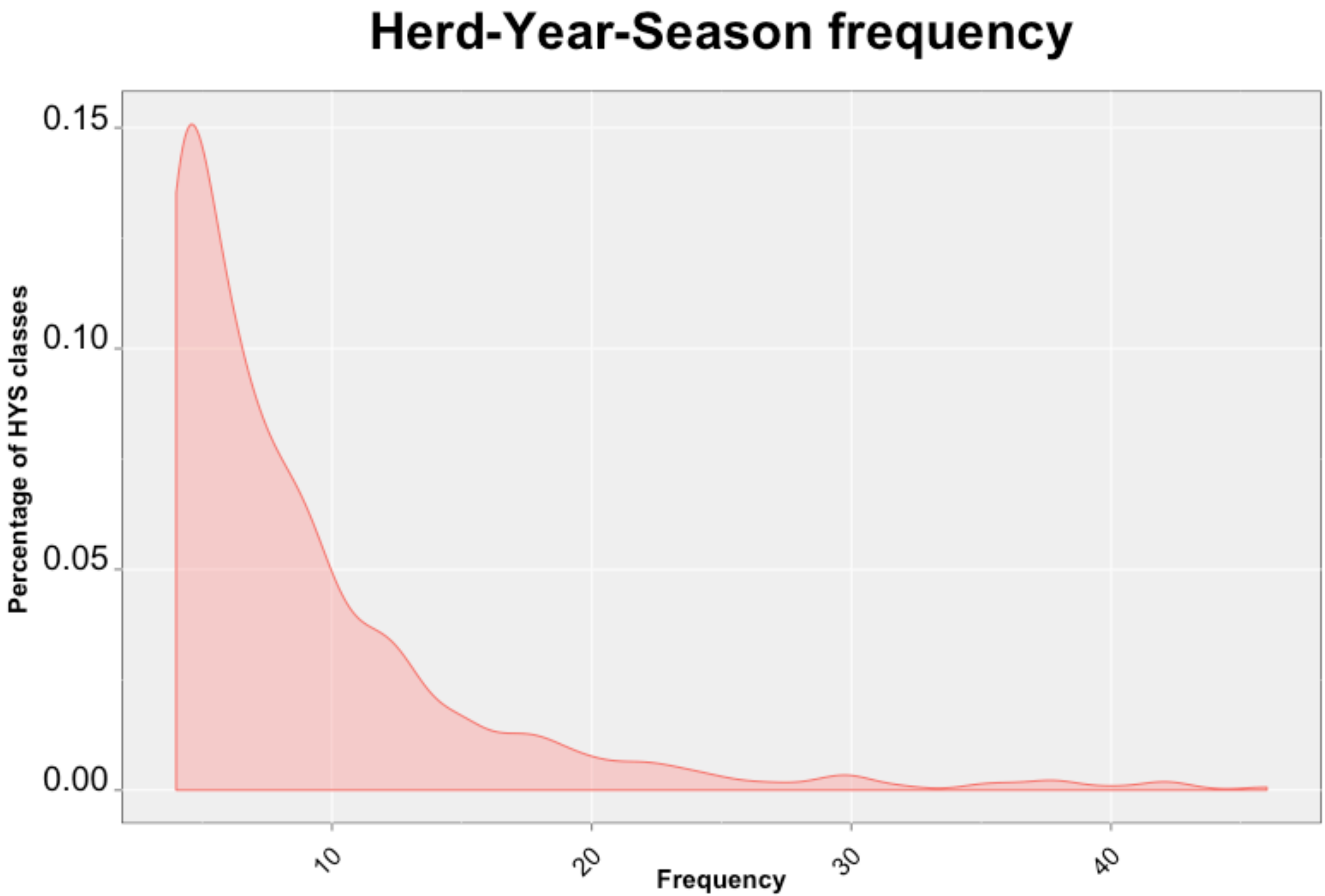
Management Data

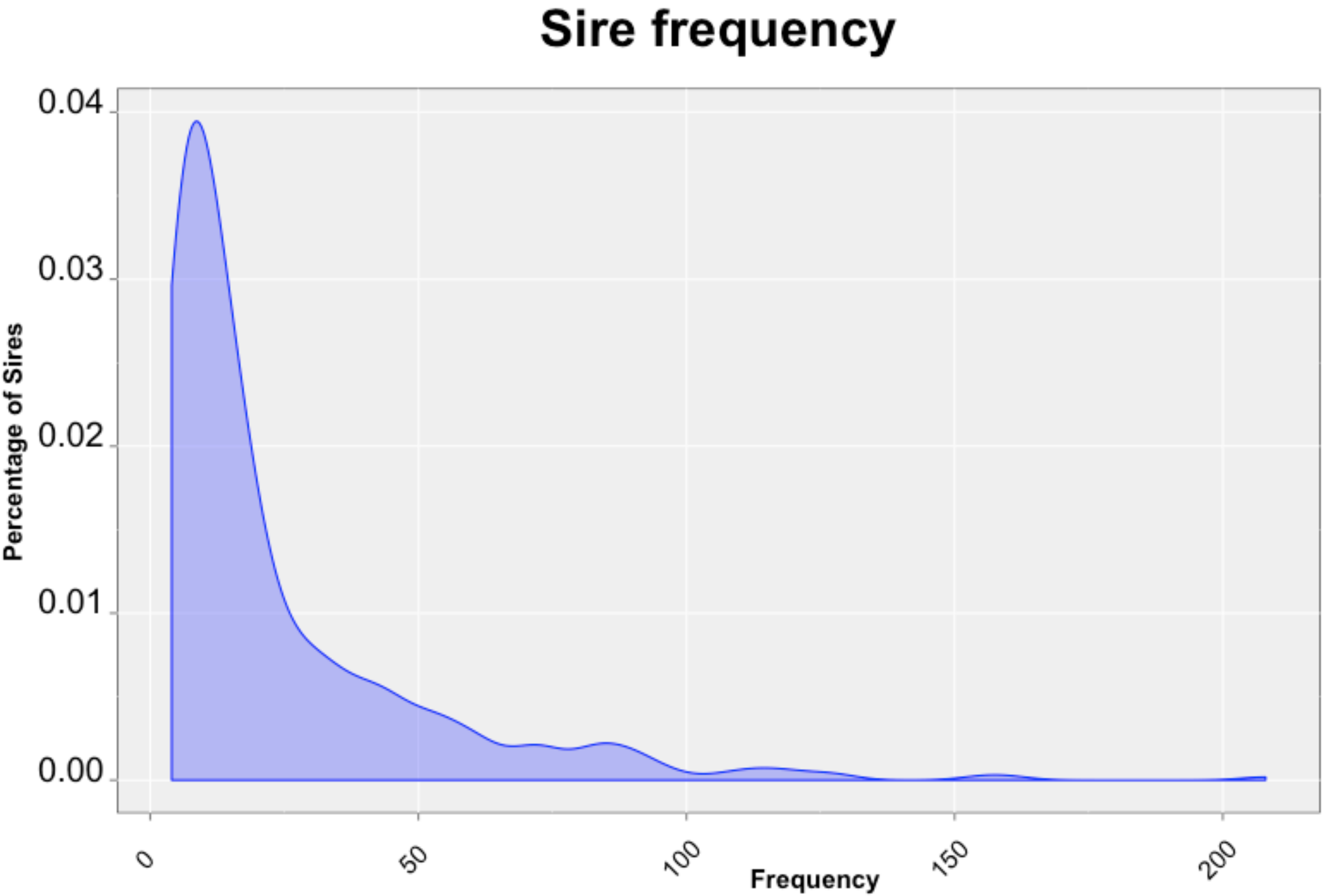
- **Herdcode** (103 levels)

- **HYS measures**
  - Number of Heads (940±654)
  - Percentage of Holstein cows (99.4±9.0)
  - Number of Milkings per day (2/3)

Distribution of Records across the Country




Herd-Year-Season frequency
0.15
0.10
0.05
0.00
Percentage of HYS classes
10
20
30
40
Frequency


Sire frequency
0.04
0.03
0.02
0.01
0.00
Percentage of Sires
0
50
100
150
200
Frequency

- **Models**

$$y_{ijk} = \mu + g_i + e_{ijk}$$

  - **G**

$$e_{ijk} \sim N(0, \mathrm{I}\sigma^2{}_{\mathrm{e}})$$
$$g_i \sim N(0, \mathrm{G}\sigma^2{}_{\mathrm{g}})$$

- **Models**

$$y_{ijk} = \mu + g_i + l_j + e_{ijk}$$

  - G
  - G+L

$$e_{ijk} \sim N(0, \mathbf{I}\sigma^2{}_e)$$
$$g_i \sim N(0, \mathbf{G}\sigma^2{}_g)$$
$$l_j \sim N(0, \mathbf{LL'}\sigma^2{}_l)$$

L { **Latitude**, **Longitude** }

- **Models**

$$y_{ijk} = \mu + g_i + l_j + gl_{ij} + e_{ijk}$$

  - G
  - G+L
  - **G+L+GxL**

$$e_{ijk} \sim N(0, \mathbf{I}\sigma^2_e)$$
$$g_i \sim N(0, \mathbf{G}\sigma^2_g)$$
$$l_j \sim N(0, \mathbf{LL'}\sigma^2_l)$$
$$gl_{ij} \sim N(0, [\mathbf{G} \circ \mathbf{LL'}]\sigma^2_{gl})$$

L { **Latitude**, **Longitude** }

- **Models**

$$y_{ijk} = \mu + g_i + w_j + e_{ijk}$$

  - G
  - G+L
  - G+L+GxL
  - **G+W**

$$e_{ijk} \sim N(0, \mathrm{I}\sigma^2_e)$$
$$g_i \sim N(0, \mathrm{G}\sigma^2_g)$$
$$w_j \sim N(0, \mathrm{WW}'\sigma^2_w)$$

W {
Maximum Temperature
Minimum Temperature
Average Temperature
Relative Humidity
Pressure
Wind Speed
Rainfall

- **Models**

$$y_{ijk} = \mu + g_i + w_j + gw_{ij} + e_{ijk}$$

- G
- G+L
- G+L+GxL
- G+W
- **G+W+GxW**

$$e_{ijk} \sim N(0, \mathbf{I}\sigma^2_e)$$
$$g_i \sim N(0, \mathbf{G}\sigma^2_g)$$
$$w_j \sim N(0, \mathbf{WW'}\sigma^2_w)$$
$$gw_{ij} \sim N(0, [\mathbf{G} \circ \mathbf{WW'}]\sigma^2_{gw})$$

W {
Maximum Temperature
Minimum Temperature
Average Temperature
Relative Humidity
Pressure
Wind Speed
Rainfall

- **Models**

$$y_{ijk} = \mu + g_i + m_j + e_{ijk}$$

- G
- G+L
- G+L+GxL
- G+W
- G+W+GxW
- **G+M**

$$e_{ijk} \sim N(0, \mathbf{I}\sigma^2_e)$$
$$g_i \sim N(0, \mathbf{G}\sigma^2_g)$$
$$m_j \sim N(0, \mathbf{MM'}\sigma^2_m)$$

M { **Number of Heads**, **Percentage HO**, **Milkings per Day** }

- **Models**

$$y_{ijk} = \mu + g_i + m_j + gm_{ij} + e_{ijk}$$

  - G
  - G+L
  - G+L+GxL
  - G+W
  - G+W+GxW
  - G+M
  - **G+M+GxM**

$$e_{ijk} \sim N(0, \mathbf{I}\sigma^2_e)$$
$$g_i \sim N(0, \mathbf{G}\sigma^2_g)$$
$$m_j \sim N(0, \mathbf{MM'}\sigma^2_m)$$
$$gm_{ij} \sim N(0, [\mathbf{G} \circ \mathbf{MM'}]\sigma^2_{gm})$$

M {
- **Number of Heads**
- **Percentage HO**
- **Milkings per Day**

- **Models**

$$y_{ijk} = \mu + g_i + h_j + e_{ijk}$$

$$e_{ijk} \sim N(0, \mathrm{I}\sigma^2_e)$$
$$g_i \sim N(0, \mathrm{G}\sigma^2_g)$$
$$h_j \sim N(0, \mathrm{HH'}\sigma^2_h)$$

  - G
  - G+L
  - G+L+GxL
  - G+W
  - G+W+GxW
  - G+M
  - G+M+GxM
  - **G+H**

H { **Herdcodes**

- **Models**

$$y_{ijk} = \mu + g_i + h_j + gh_{ij} + e_{ijk}$$

- G
- G+L
- G+L+GxL
- G+W
- G+W+GxW
- G+M
- G+M+GxM
- G+H
- **G+H+GxH**

$$e_{ijk} \sim N(0, \mathbf{I}\sigma^2_e)$$
$$g_i \sim N(0, \mathbf{G}\sigma^2_g)$$
$$h_j \sim N(0, \mathbf{HH'}\sigma^2_h)$$
$$gh_{ij} \sim N(0, [\mathbf{G} \circ \mathbf{HH'}]\sigma^2_{gh})$$

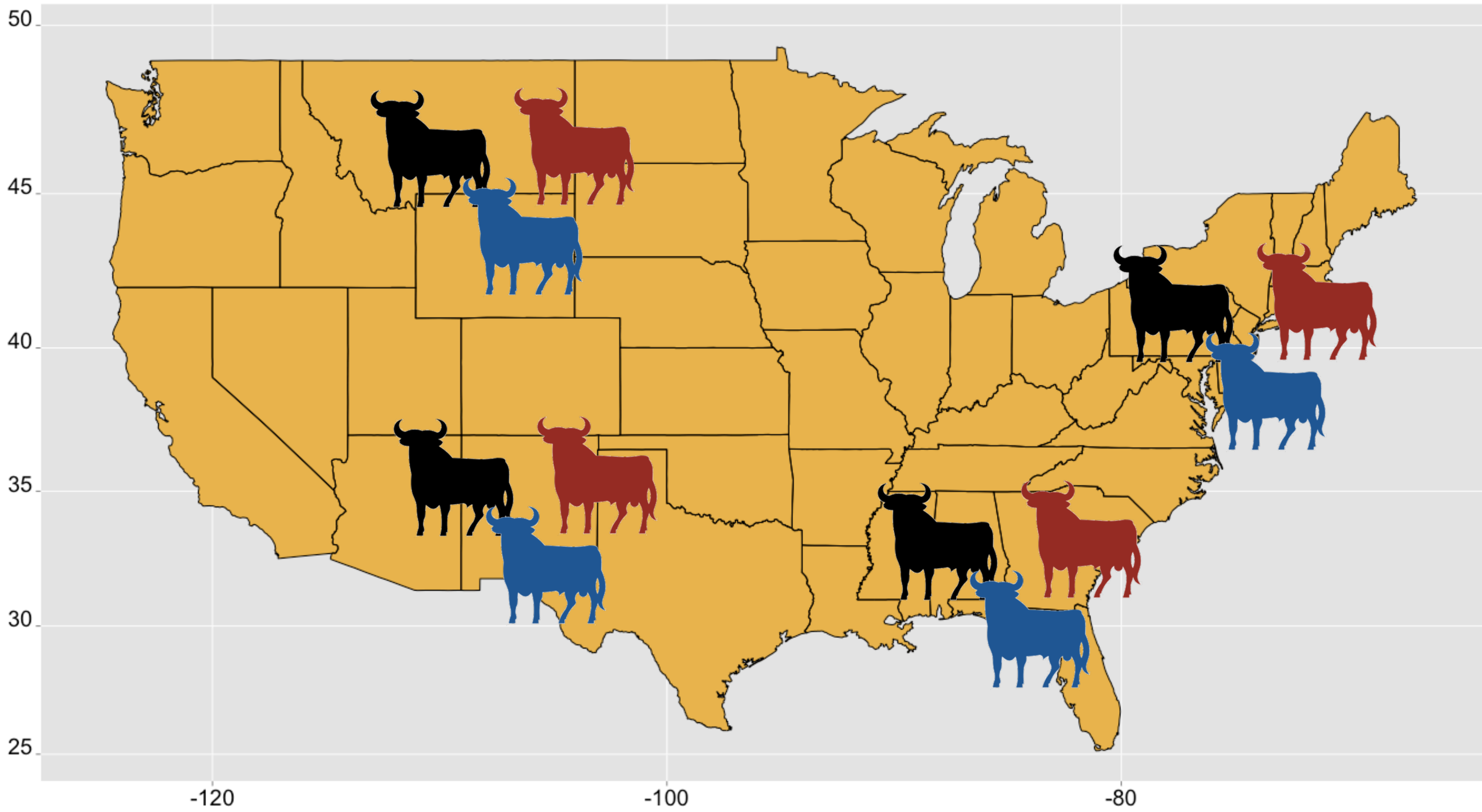
H { **Herdcodes**

- **Models**

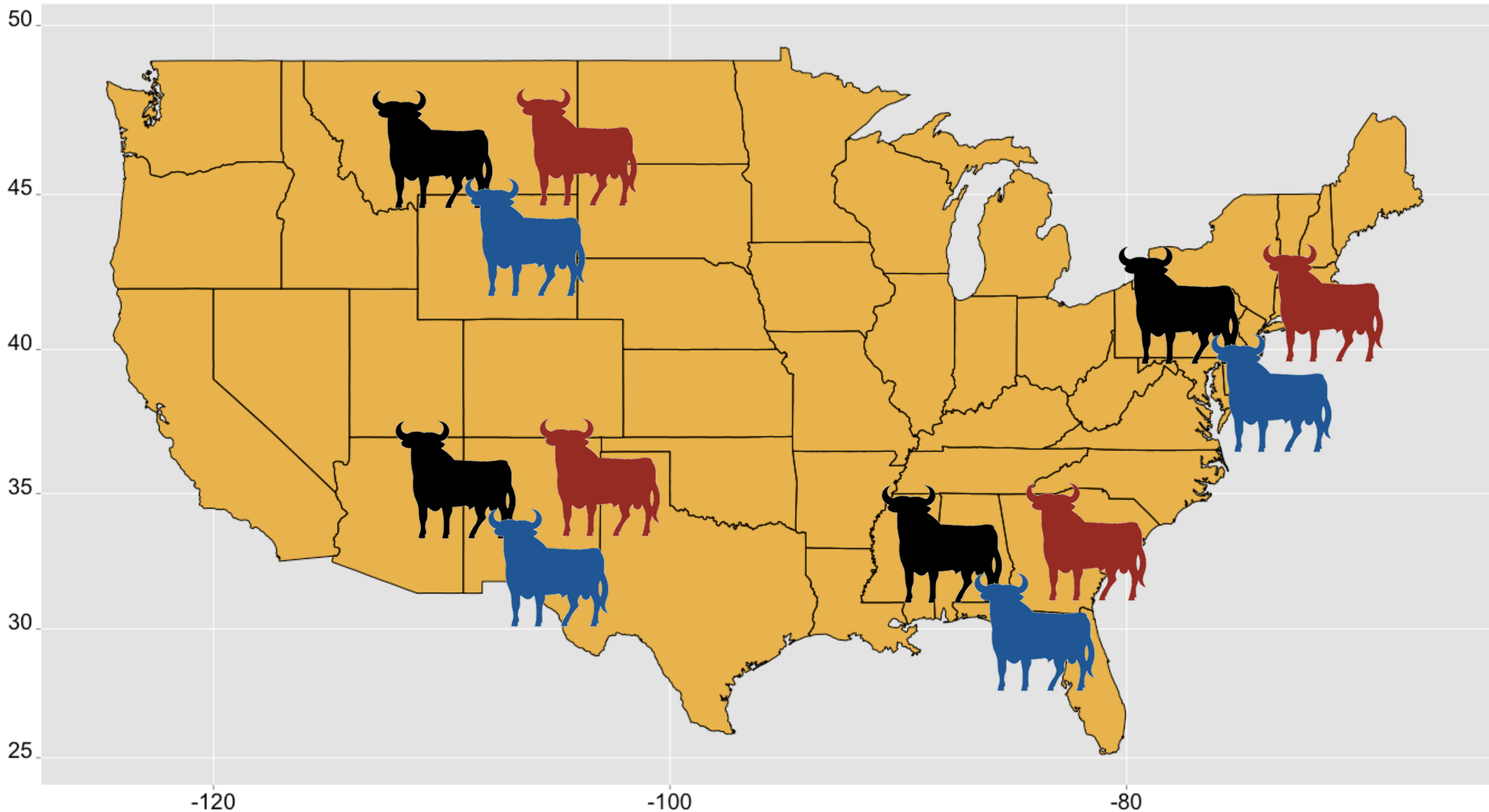
  - **BGLR** **(Perez & de los Campos, 2014; Jarquin et al., 2014)**
  - **Eigenvalue decomposition** **(Janss et al., 2012)**

  - **62,000 iterations**
    - **2,000 as burn-in**
    - **thin every 10**

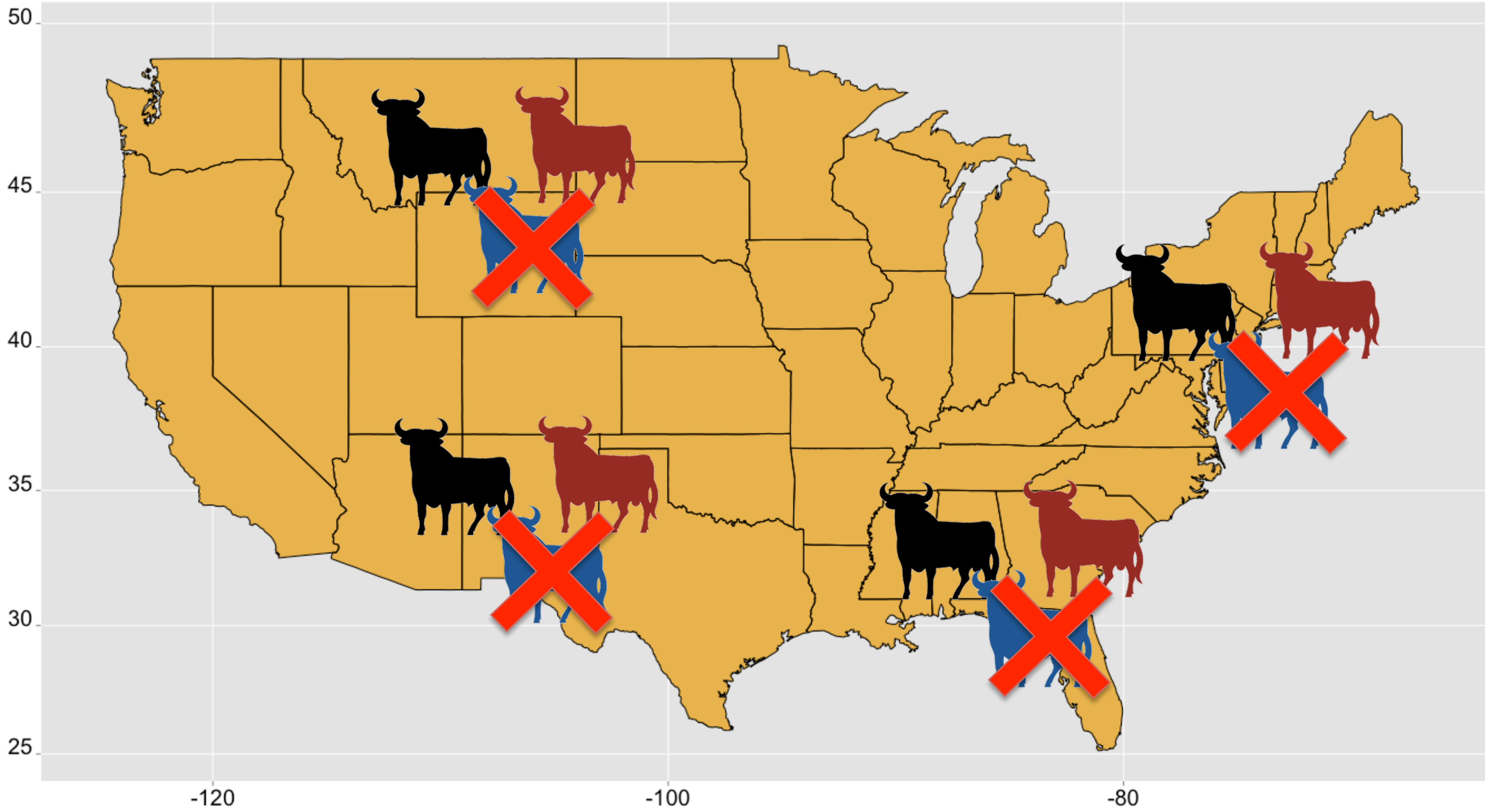
Cross Validation



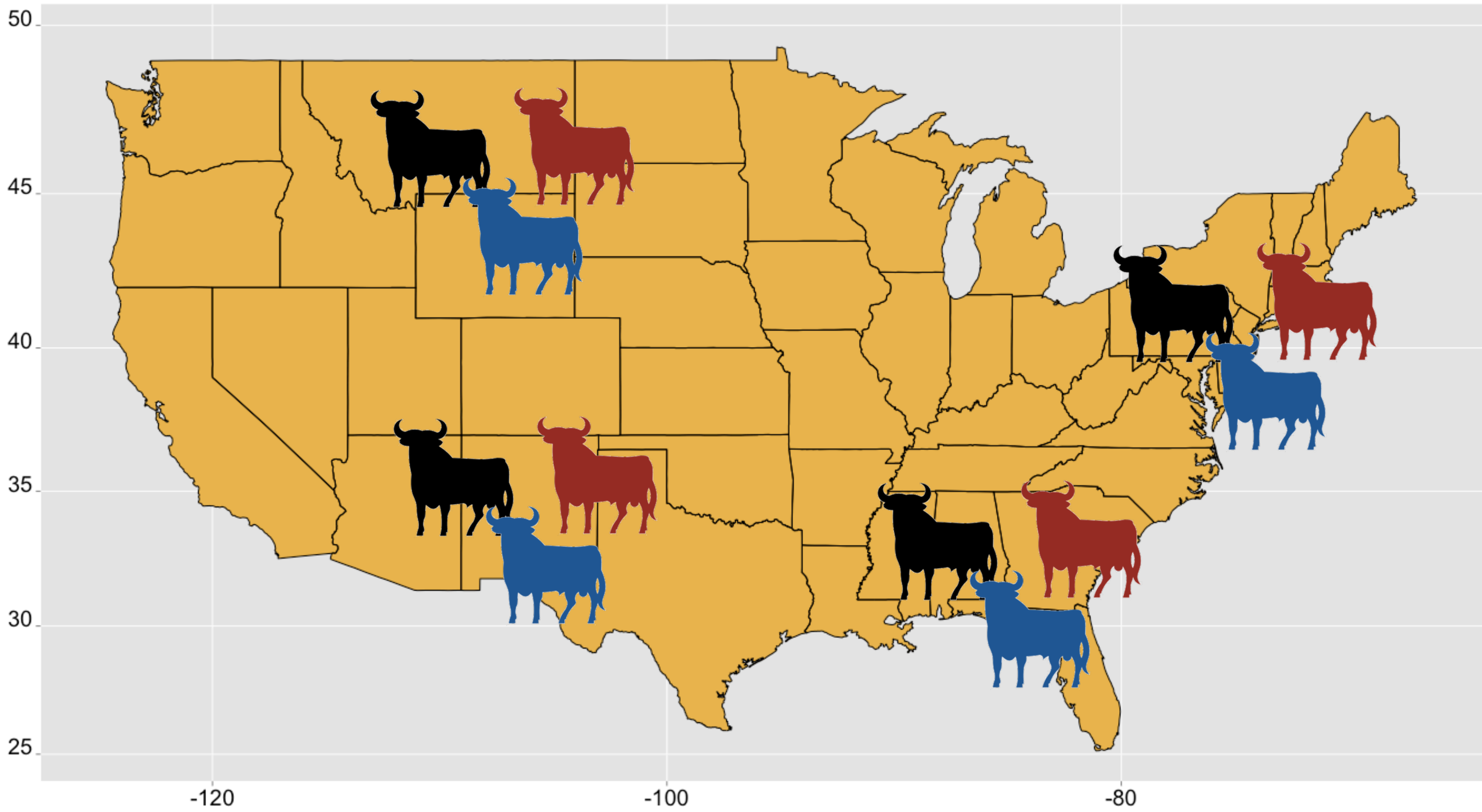
CV: Split 1 - New Bulls



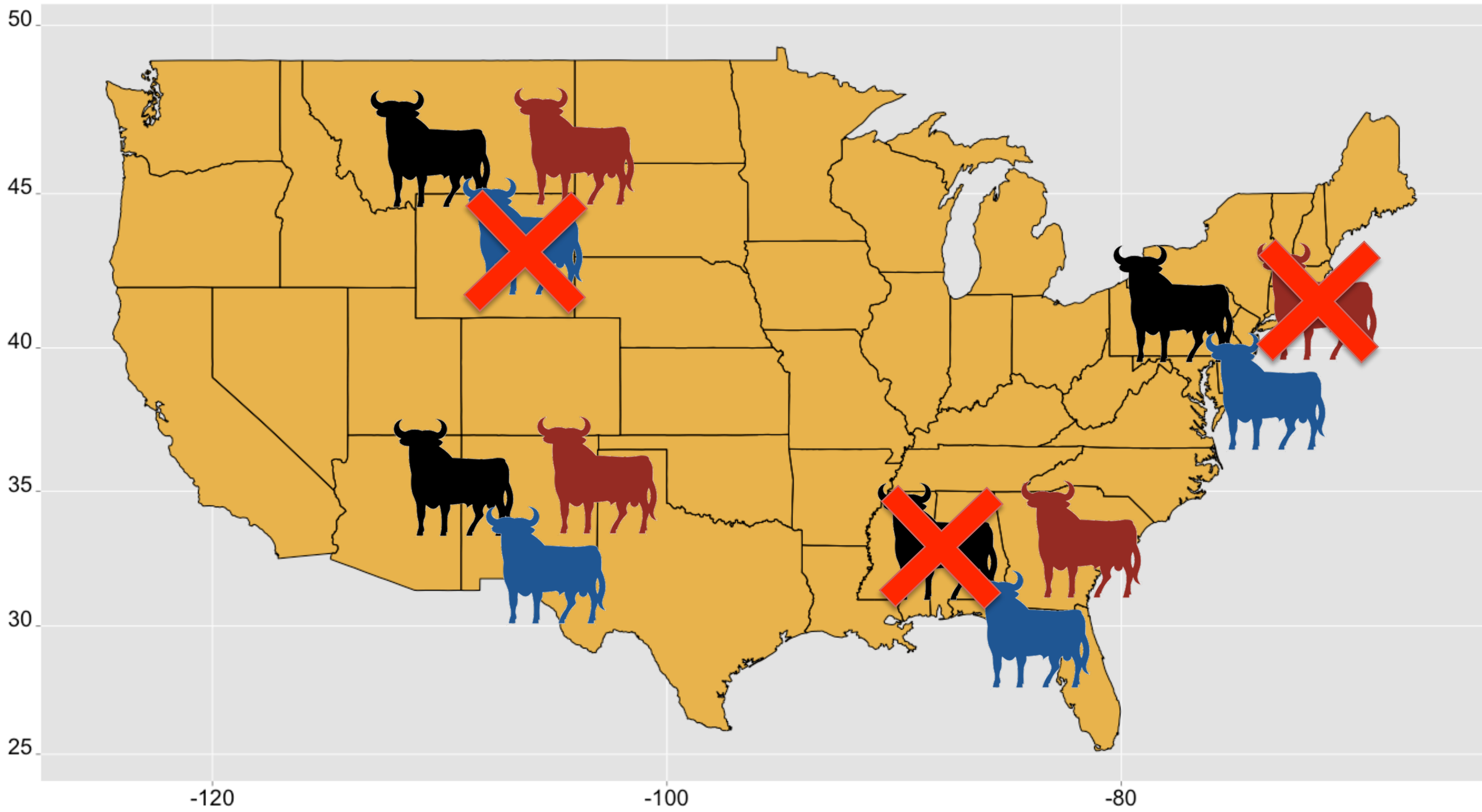
CV: Split 1 - New Bulls



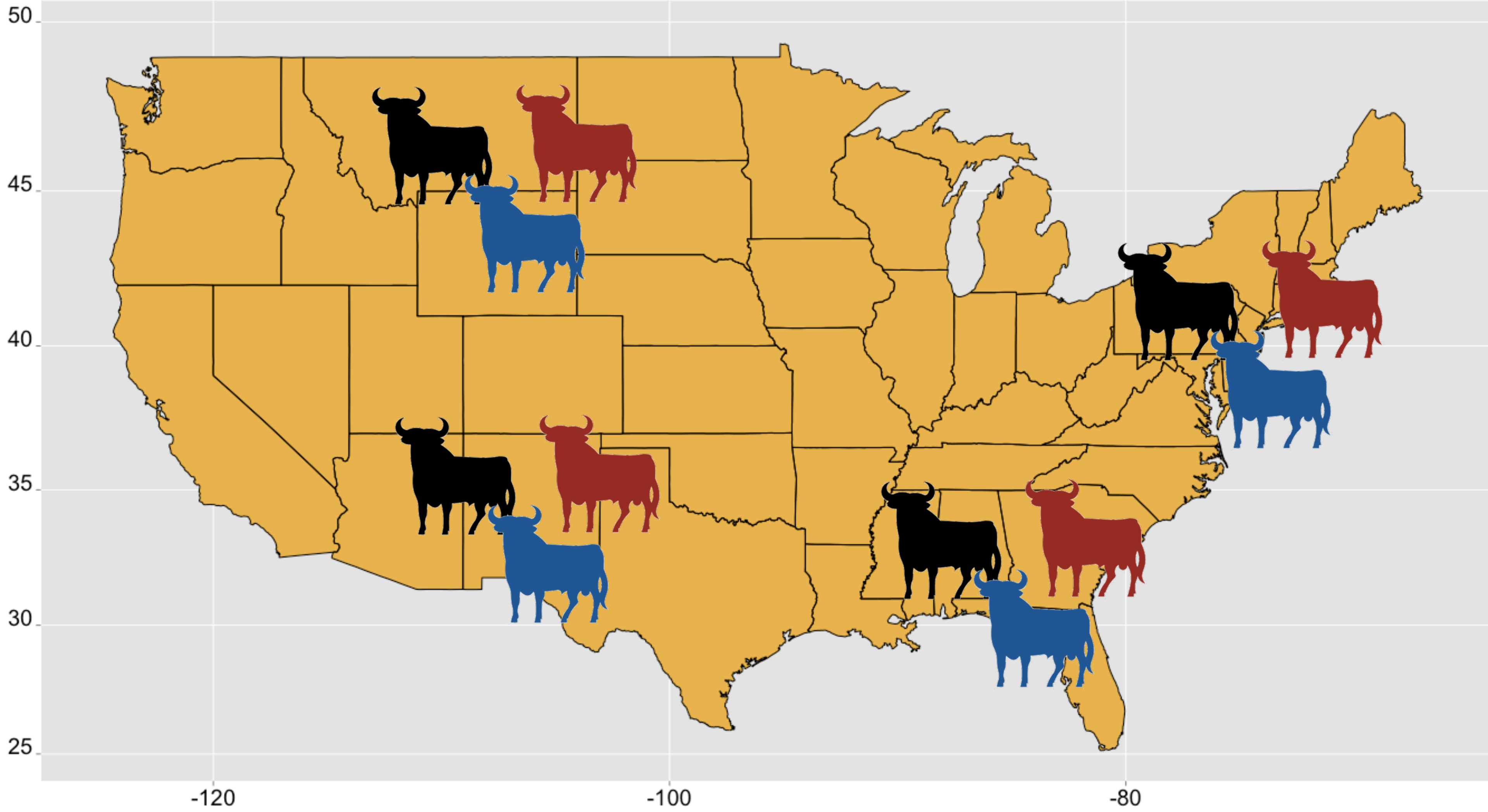
CV: Split 2 - Incomplete Progeny Test



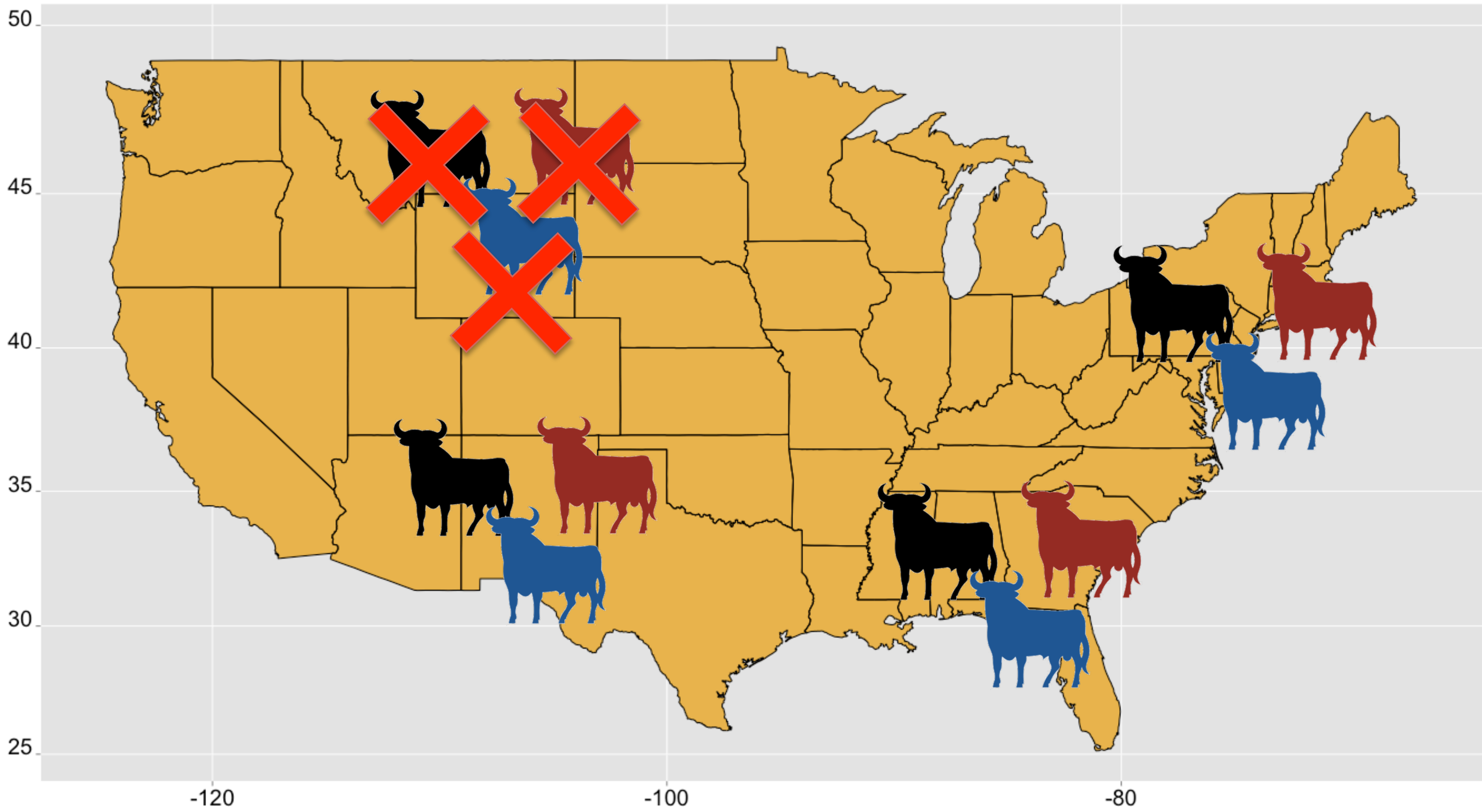
CV: Split 2 - Incomplete Progeny Test



CV: Split 3 - Missing Region



CV: Split 3 - Missing Region



	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
New bulls	1,021 (9%)	1,039 (9%)	1,905 (16%)	1,373 (12%)	.
Inc. Prog.	522 (5%)	575 (5%)	668 (6%)	663 (6%)	.
Missing Region	2,730 (23%)	1,757 (15%)	3,064 (26%)	1,152 (10%)	1,993 (17%)

Accuracy -> $Corr(y,\hat{y})$

Results

Variance Components Estimates

Model	G	L	GxL	W	GxW	M	GxM	H	GxH
G	0.568	.	.	.	.	.	.	.	.
G+L	0.521	0.054	.	.	.	.	.	.	.
G+L+GxL	0.192	0.062	0.376	.	.	.	.	.	.
G+W	0.526	.	.	0.098	.	.	.	.	.
G+W+GxW	0.382	.	.	0.061	0.168	.	.	.	.
G+M	0.449	.	.	.	.	0.037	.	.	.
G+M+GxM	0.277	.	.	.	.	0.036	0.283	.	.
G+H	0.372	.	.	.	.	.	.	0.354	.
G+H+GxH	0.372	.	.	.	.	.	.	0.352	0.004

Variance Components Estimates

Model	G	L	GxL	W	GxW	M	GxM	H	GxH
G	0.568	.	.	.	.	.	.	.	.
G+L	0.521	0.054	.	.	.	.	.	.	.
G+L+GxL	0.192	0.062	0.376	.	.	.	.	.	.
G+W	0.526	.	.	0.098	.	.	.	.	.
G+W+GxW	0.382	.	.	0.061	0.168	.	.	.	.
G+M	0.449	.	.	.	.	0.037	.	.	.
G+M+GxM	0.277	.	.	.	.	0.036	0.283	.	.
G+H	0.372	.	.	.	.	.	.	0.354	.
G+H+GxH	0.372	.	.	.	.	.	.	0.352	0.004

Variance Components Estimates

Model	G	L	GxL	W	GxW	M	GxM	H	GxH
G	0.568	.	.	.	.	.	.	.	.
G+L	0.521	0.054	.	.	.	.	.	.	.
G+L+GxL	0.192	0.062	0.376	.	.	.	.	.	.
G+W	0.526	.	.	0.098	.	.	.	.	.
G+W+GxW	0.382	.	.	0.061	0.168	.	.	.	.
G+M	0.449	.	.	.	.	0.037	.	.	.
G+M+GxM	0.277	.	.	.	.	0.036	0.283	.	.
G+H	0.372	.	.	.	.	.	.	0.354	.
G+H+GxH	0.372	.	.	.	.	.	.	0.352	0.004

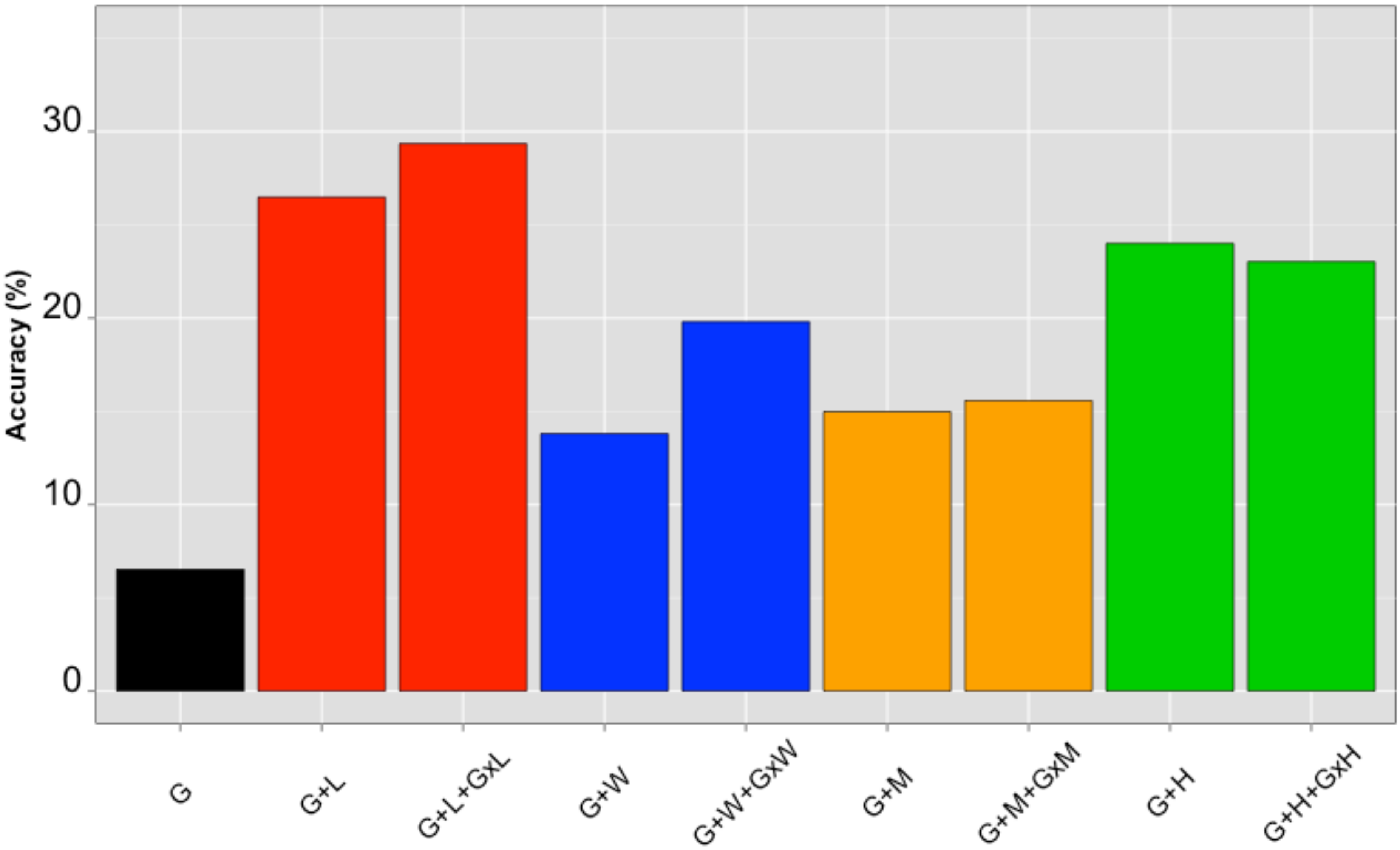
Variance Components Estimates

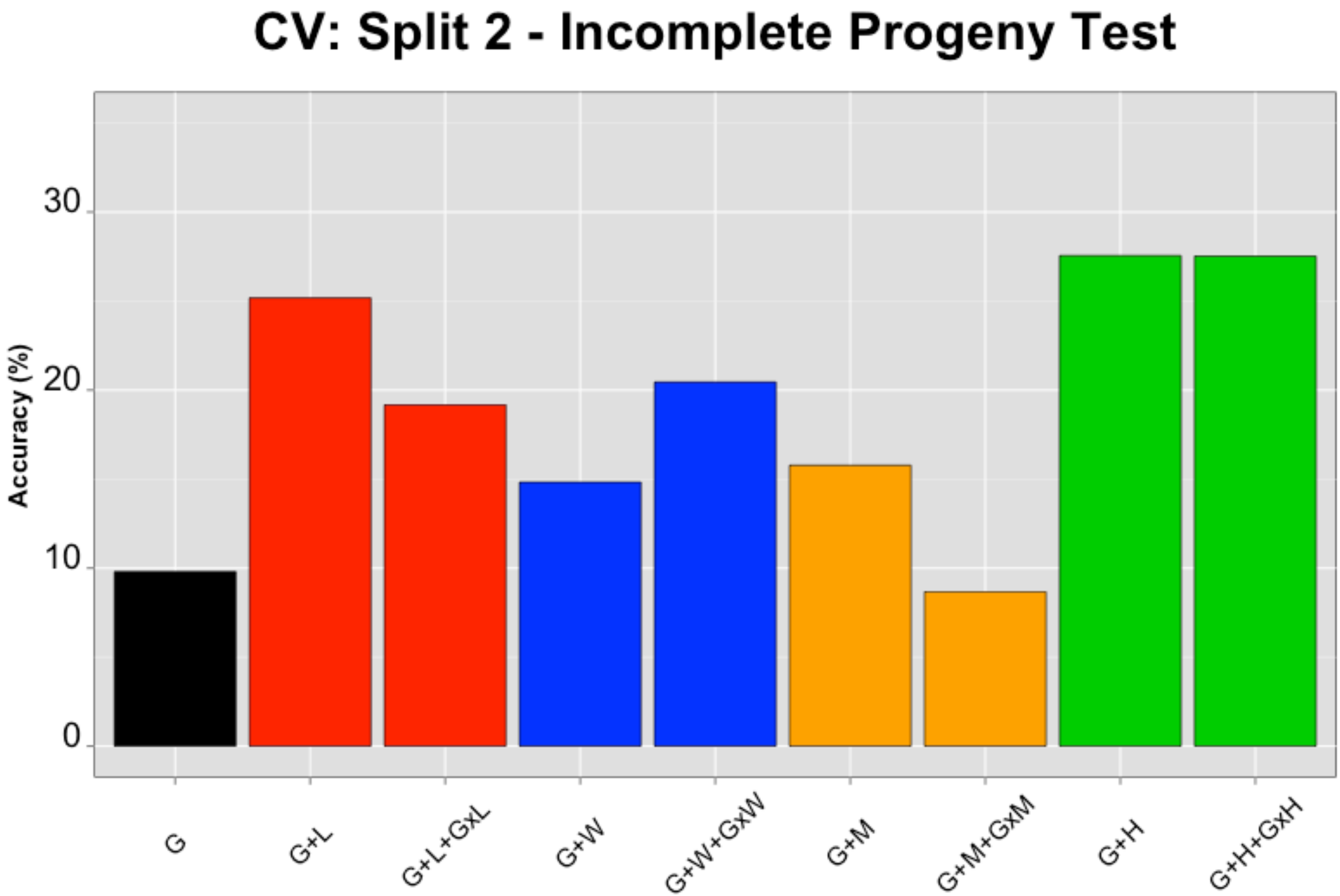
Model	G	L	GxL	W	GxW	M	GxM	H	GxH
G	0.568	.	.	.	.	.	.	.	.
G+L	0.521	0.054	.	.	.	.	.	.	.
G+L+GxL	0.192	0.062	0.376	.	.	.	.	.	.
G+W	0.526	.	.	0.098	.	.	.	.	.
G+W+GxW	0.382	.	.	0.061	0.168	.	.	.	.
G+M	0.449	.	.	.	.	0.037	.	.	.
G+M+GxM	0.277	.	.	.	.	0.036	0.283	.	.
G+H	0.372	.	.	.	.	.	.	0.354	.
G+H+GxH	0.372	.	.	.	.	.	.	0.352	0.004

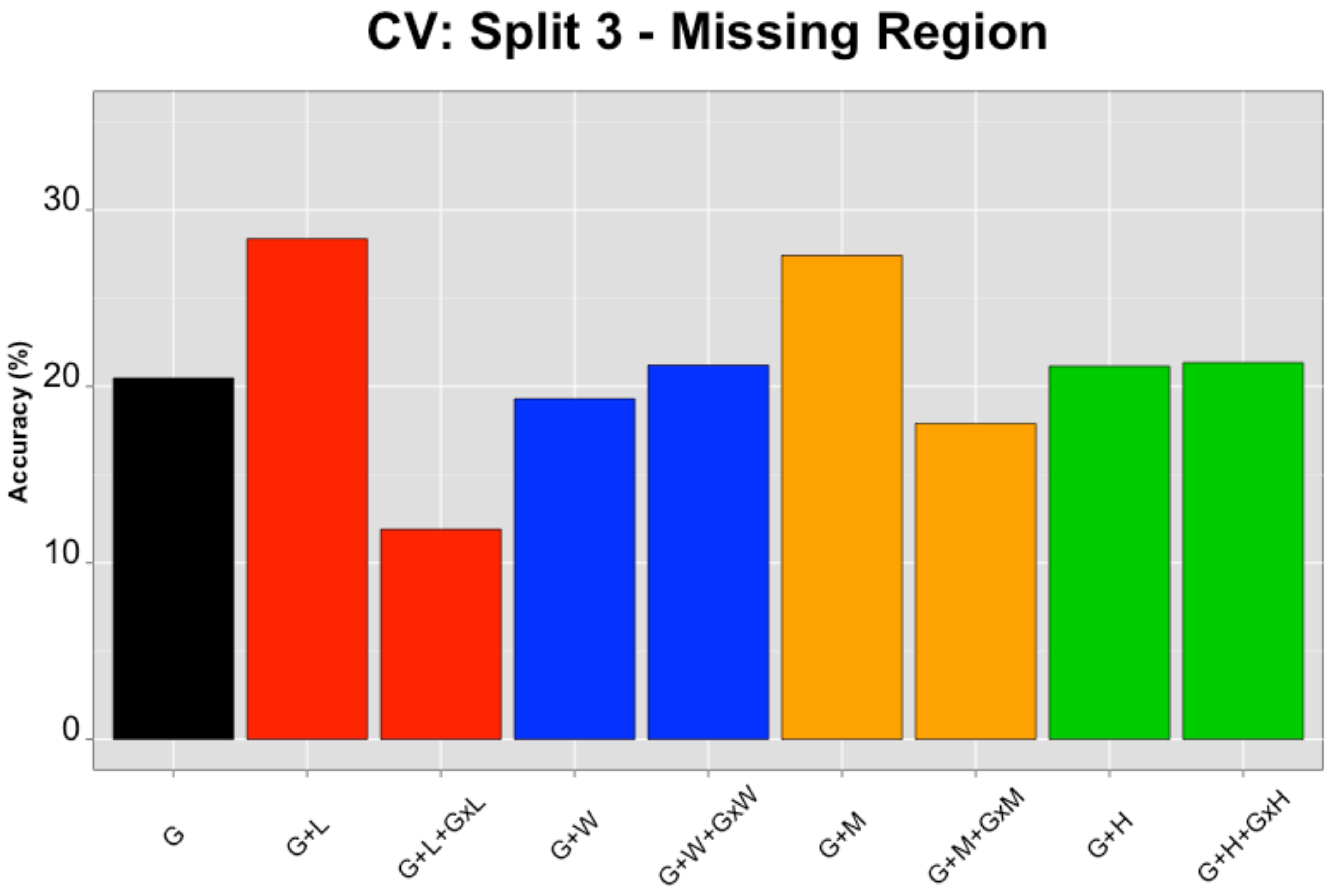
Variance Components Estimates

Model	G	L	GxL	W	GxW	M	GxM	H	GxH
G	0.568	.	.	.	.	.	.	.	.
G+L	0.521	0.054	.	.	.	.	.	.	.
G+L+GxL	0.192	0.062	0.376	.	.	.	.	.	.
G+W	0.526	.	.	0.098	.	.	.	.	.
G+W+GxW	0.382	.	.	0.061	0.168	.	.	.	.
G+M	0.449	.	.	.	.	0.037	.	.	.
G+M+GxM	0.277	.	.	.	.	0.036	0.283	.	.
G+H	0.372	.	.	.	.	.	.	0.354	.
G+H+GxH	0.372	.	.	.	.	.	.	0.352	0.004

CV: Split 1 - New Bulls




CV: Split 2 - Incomplete Progeny Test
Accuracy (%)
0
10
20
30
G
G+L
G+L+GxL
G+W
G+W+GxW
G+M
G+M+GxM
G+H
G+H+GxH


CV: Split 3 - Missing Region
Accuracy (%)
30
20
10
0
G
G+L
G+L+GxL
G+W
G+W+GxW
G+M
G+M+GxM
G+H
G+H+GxH

Conclusions and Future Perspectives

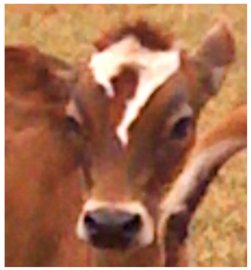
- **GxE exists, but not always estimable**
- **Not a single set of covariables that predicts well accounting for GxE**
- **Climate covariables are the only to yield better predictions with GxE**

→ Find better combinations of predictors ←

- **Repeat analysis on other traits**
- **Increase number of genotyped bulls**


Thank you for your attention!
NC STATE UNIVERSITY
Animal Science Department
Genetics and Genomics

Questions



Thank you for your attention!