

Report on the weaning weight phenotypic data editing

Mohammad A. Nilforooshan

Mohammad.nilforooshan@slu.se

Interbeef technical committee meeting

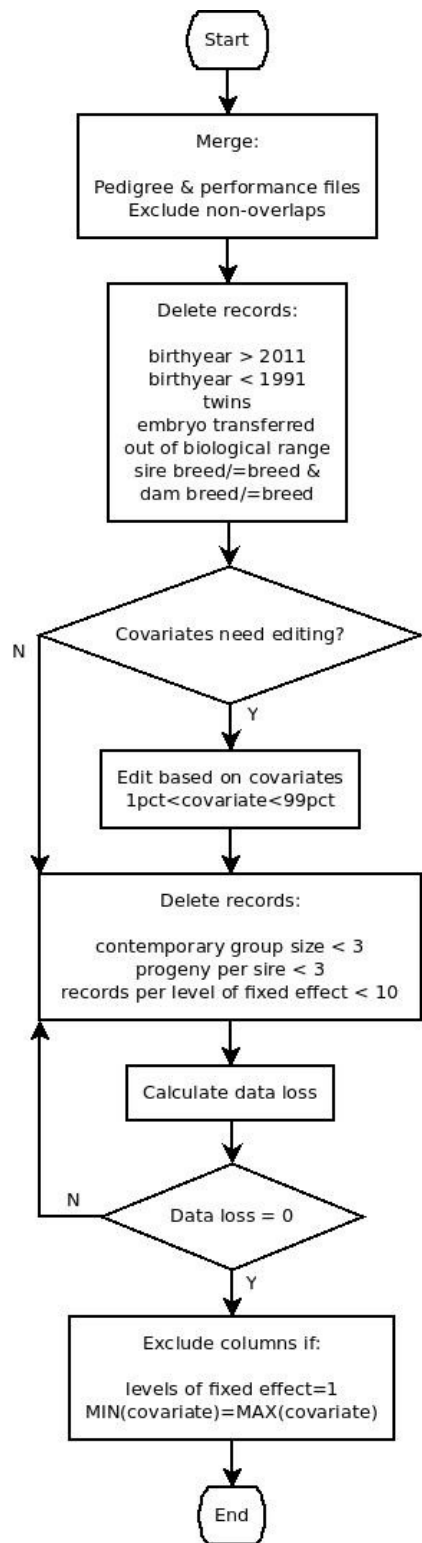
Uppsala, December 6-7, 2012

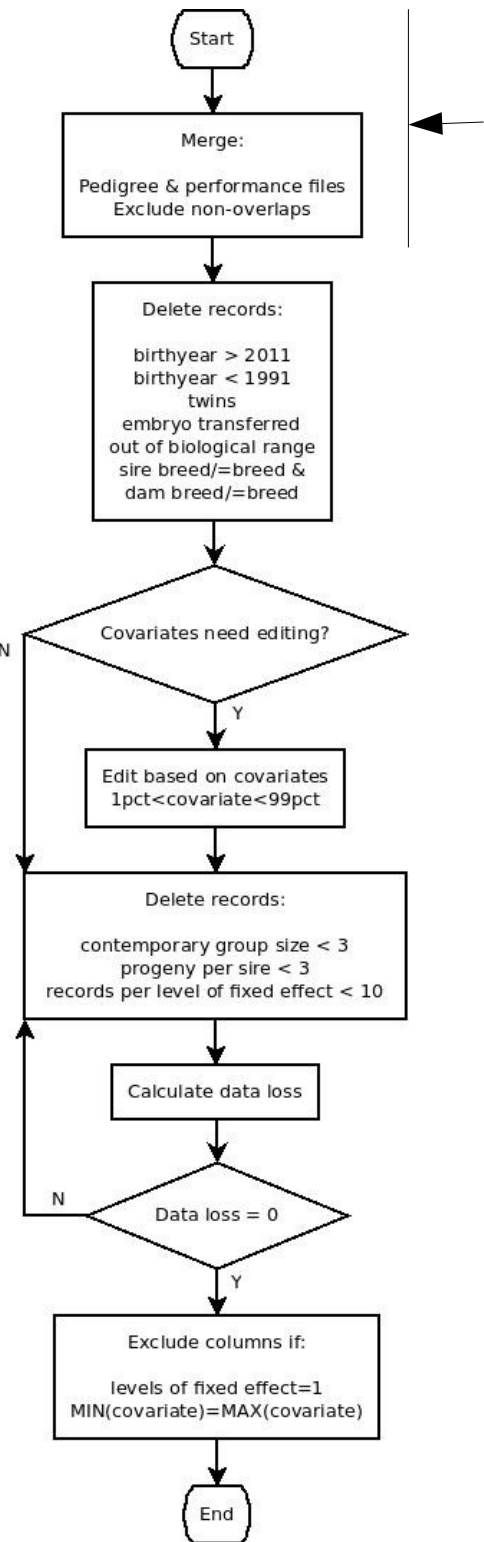
Why data editing?

good input + good processing = good output



Data editing flowchart





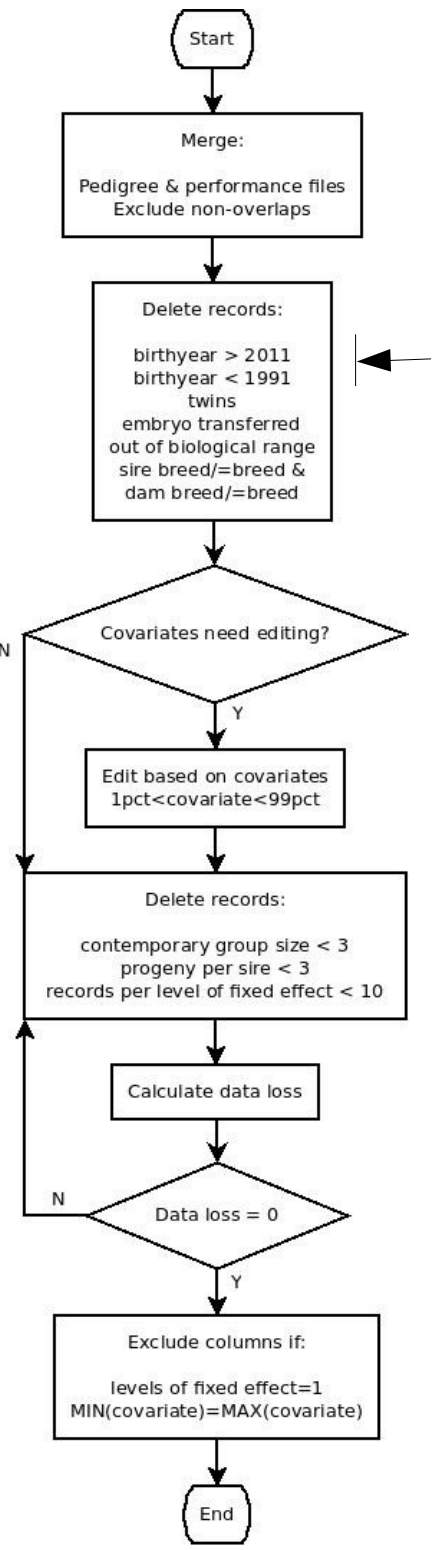
Merge:
PED & PERF
Exclude non-overlaps

Data loss after merging PED & PERF CHA:

CZE	0	24,371
DNK	0	15,564
FIN	1	17,185
FRA	2	3,472,920
SWE	1	99,691
IRL	2,776	20,216

Data loss after merging PED & PERF LIM:

CZE	0	5,796
DNK	0	40,737
ESP	0	33,259
FIN	0	13,036
FRA	23	2,083,507
GBR	4,226	112,845
SWE	0	18,201
IRL	1,031	18,806



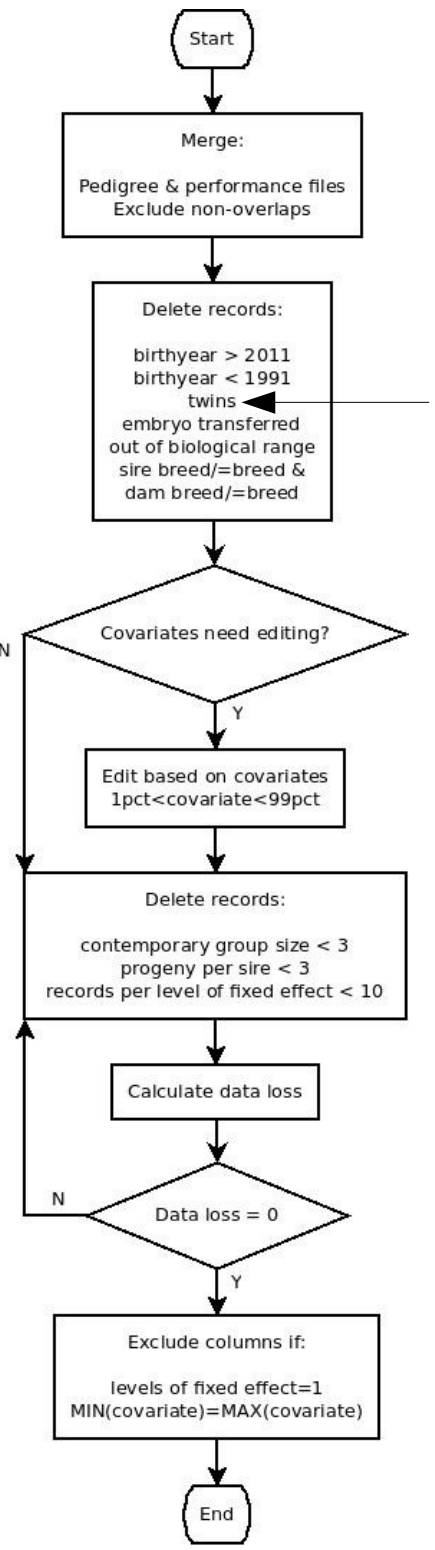
Keep:
1990 < byear < 2012

Filter data by byear, CHA:

	< 1991	1990<byear<2012	> 2011
CZE	0	18,923	5
DNK	2,600	12,347	0
FIN	660	15,305	14
FRA	598,854	2,810,416	0
SWE	10,014	89,085	0
IRL	469	17,802	0

Filter data by byear, LIM:

	< 1991	1990<byear<2012	> 2011
CZE	0	5,070	0
DNK	6,170	33,087	11
ESP	32	32,756	0
FIN	399	11,661	8
FRA	342,968	1,680,356	0
GBR	18,988	89,741	19
SWE	1,275	16,363	0
IRL	227	17,550	0



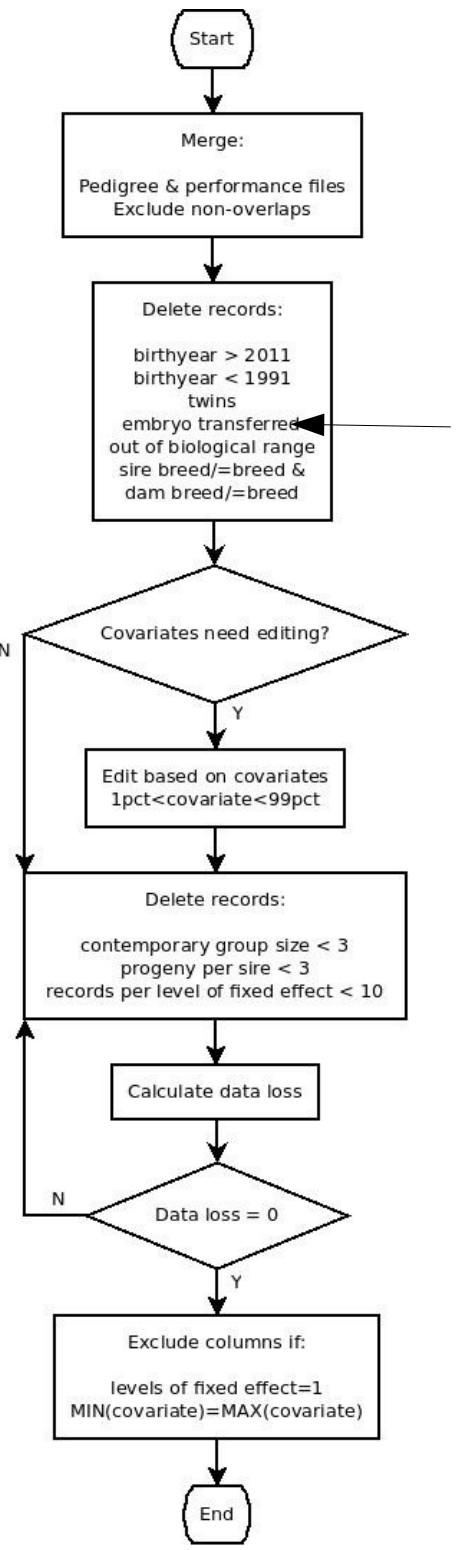
Delete:
Twins

Deleting twins, CHA:

CZE	943	17,985
DNK	880	14,067
FIN	1,100	14,879
FRA	0	3,409,270
SWE	8,411	90,688
IRL	1,283	16,988

Deleting twins, LIM:

CZE	73	4,997
DNK	633	38,635
ESP	0	32,788
FIN	164	11,904
FRA	0	2,023,324
GBR	1,835	106,913
SWE	241	17,397
IRL	553	17,224



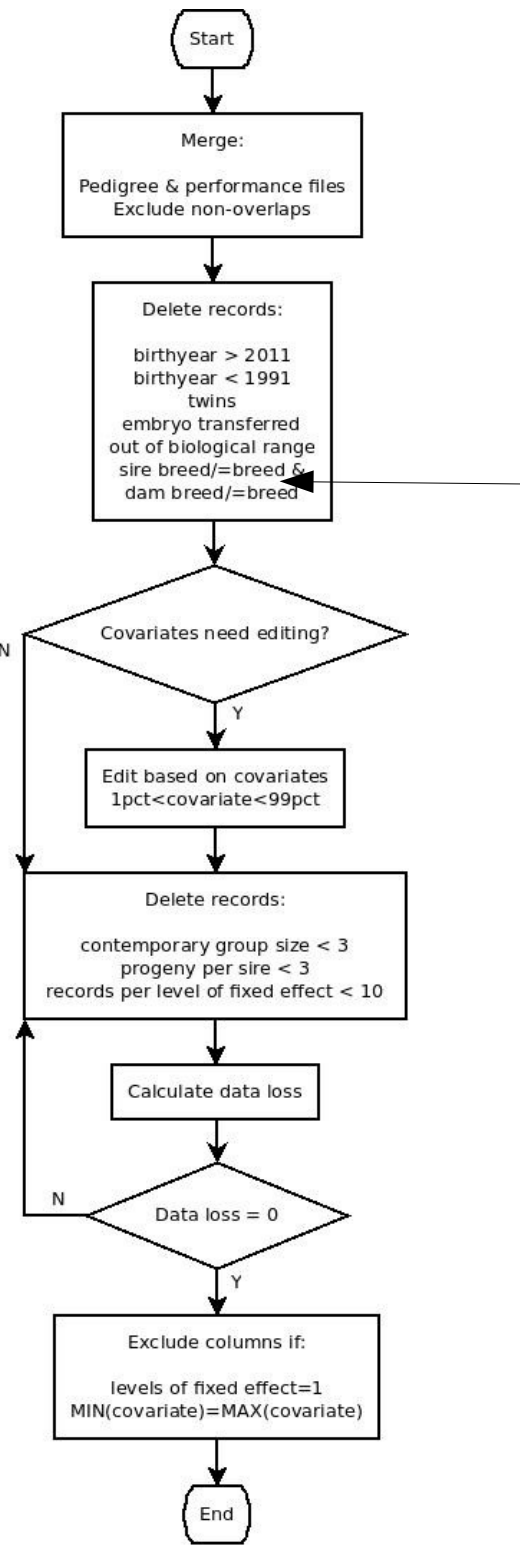
Delete:
Embryo transterred

Deleting embryo transferred

CZE	879	18,049
DNK	461	14,486
FIN	118	15,861
FRA	0	3,409,270
SWE	0	99,099
IRL	774	17,497

Deleting embryo transferred, LIM:

CZE	357	4,713
DNK	375	38,893
ESP	0	32,788
FIN	68	12,000
FRA	0	2,023,324
GBR	5,113	103,635
SWE	0	17,638
IRL	1,101	16,676



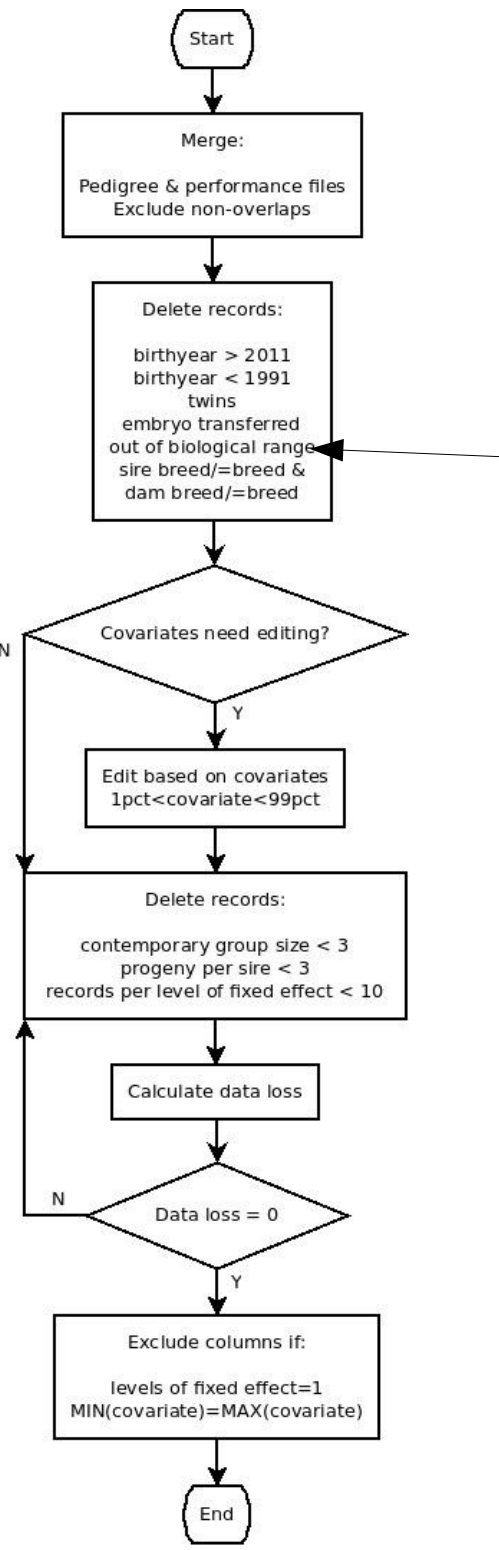
Edit:
Sire & Dam breeds

Edit for Sire and Dam breeds, CHA:

CZE	18,928	5,443
DNK	14,947	617
FIN	15,979	1,209
FRA	3,409,270	63,650
SWE	99,099	592
IRL	18,271	1,945

Edit for Sire and Dam breeds, LIM:

CZE	5,070	726
DNK	39,268	1,469
ESP	32,788	471
FIN	12,968	968
FRA	2,023,324	60,183
GBR	108,748	4,097
SWE	17,638	563
IRL	17,777	1,029



Edit:
Bio range
50 < weaning w < 500

Edit for Bio range, CHA:

CZE	18,928	0
DNK	14,947	0
FIN	15,974	5
FRA	3,409,248	22
SWE	99,092	7
IRL	-	-

A recent problem
reading the data

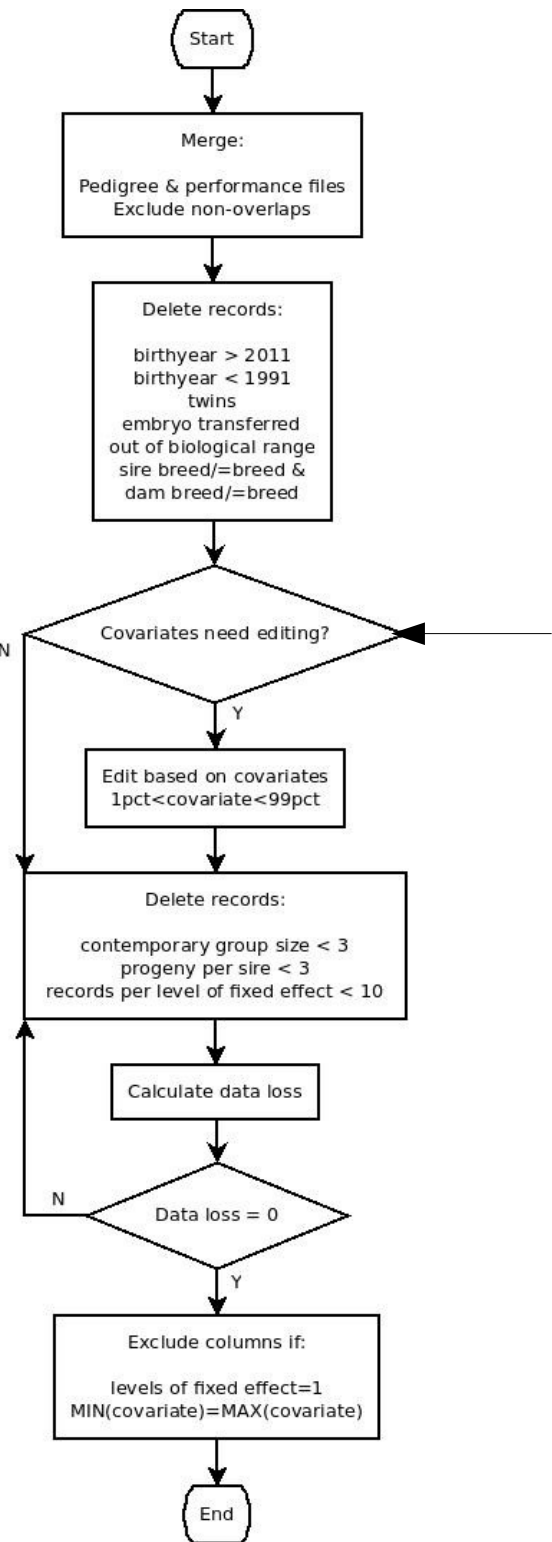
Edit for Bio range, LIM:

CZE	5,070	0
DNK	39,268	0
ESP	32,788	0
FIN	12,068	0
FRA	2,023,324	0
GBR	108,748	0
SWE	17,634	4
IRL	17,775	2

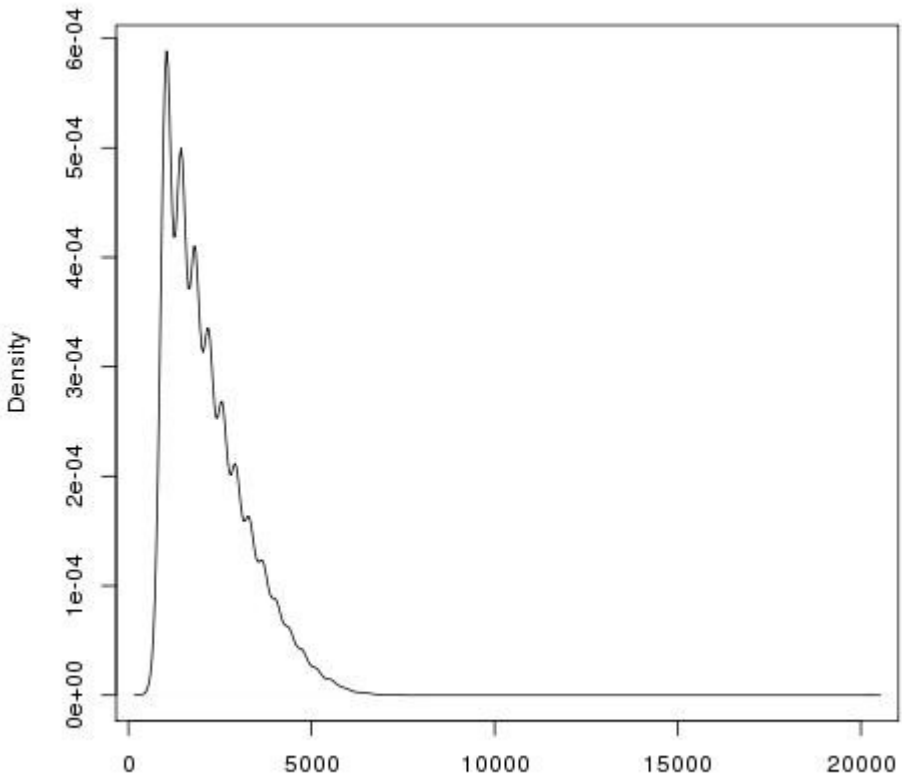
Edit for:
covariates

(age of dam) and (age of dam)² for GBR

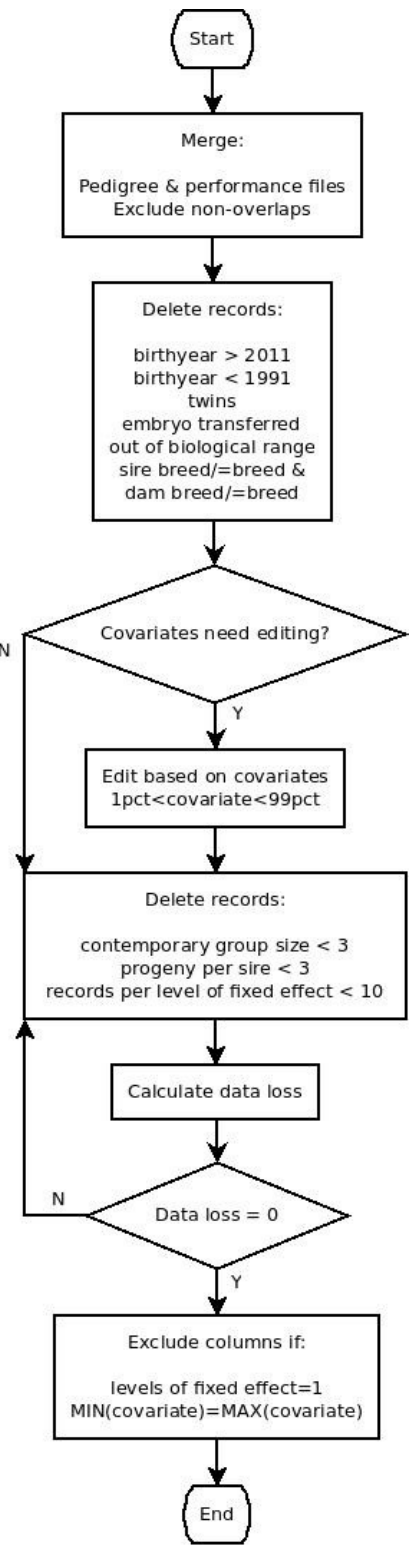
1,707 records were deleted.



Kernel Density Plot for GBR-LIM-dam_age



N = 92525 Bandwidth = 97.68



Iterative editing for:

- Contemporary group size
- Number of progeny / sire
- Number of animals in each level of a fixed effect

MIN CG size = 3, CHA:

CZE	16,216	887
DNK	10,098	1,134
FIN	13,670	482
FRA	2,870,326	30,068
SWE	79,061	2,370

MIN CG size = 3, CHA:

N Contemporary group

CZE	1,467	498
DNK	1,041	603
FIN	714	219
FRA	147,790	8,577
SWE	4,724	1,160

MIN CG size = 3, LIM:

CZE	4,321	319
DNK	30,061	2,198
ESP	32,387	369
FIN	11,136	304
FRA	1,663,839	16,517
GBR	81,483	2,062
SWE	15,582	554
IRL	13,869	2,300

MIN CG size = 3, LIM:

N Contemporary group

CZE	406	156
DNK	2,782	1,235
ESP	3,272	17
FIN	548	126
FRA	93,272	6,635
GBR	6,410	226
SWE	969	275
IRL	2,059	518

MIN progeny / sire = 3, CHA:

CZE	16,216	212
DNK	10,098	343
FIN	13,680	195
FRA	2,780,326	14,751
SWE	79,061	705

MIN progeny / sire = 3, CHA:

N Sires

CZE	550	156
DNK	552	304
FIN	618	147
FRA	62,394	10,444
SWE	2,770	719

MIN progeny / sire = 3, LIM:

CZE	4,321	101
DNK	30,061	565
ESP	32,387	346
FIN	11,136	138
FRA	1,663,839	5,654
GBR	81,493	1,795
SWE	15,582	175
IRL	13,869	2,039

MIN progeny / sire = 3, LIM:

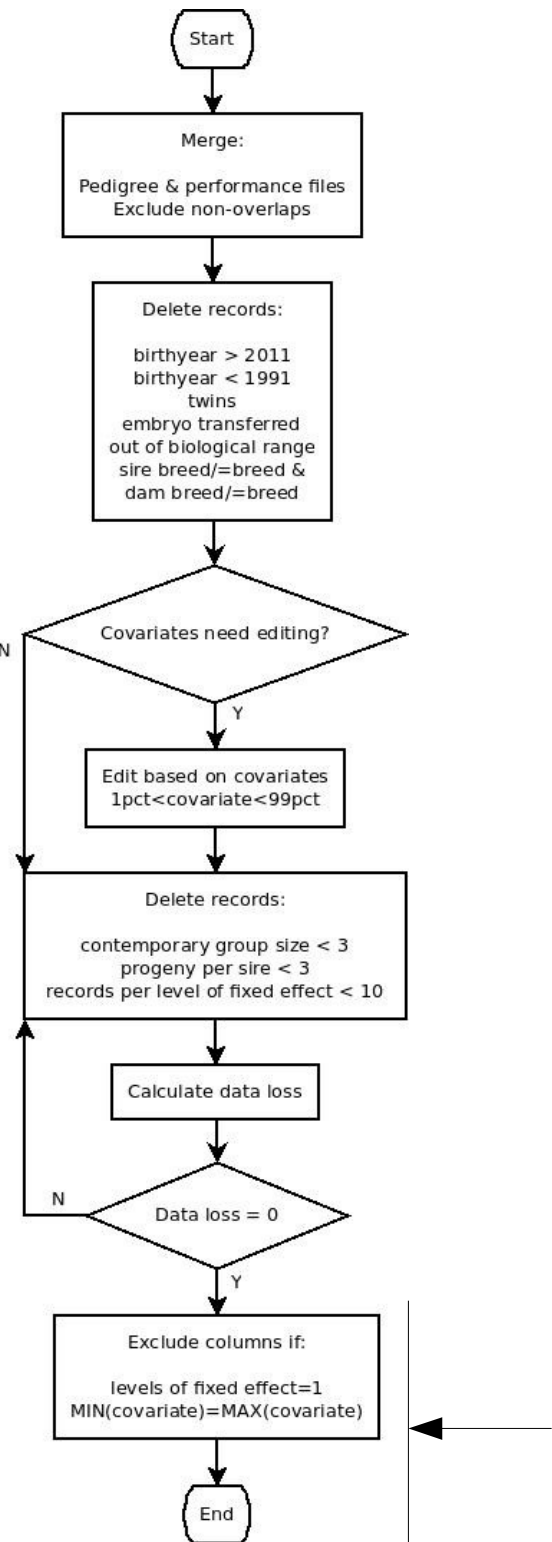
N Sires

CZE	188	83
DNK	1,357	546
ESP	925	256
FIN	450	109
FRA	31,183	4,188
GBR	2,628	1,307
SWE	634	149
IRL	719	1,278

Deleting levels of fixed effects with < 10 observations

After performing the other edits, there were not many observations deleted for this reason

(MAX)14 obs for DNK-LIM



Delete columns if:

- Nobs(level of fixed effect) < 10
- MIN(covariate) = MAX(covariate)

Conclusions

- Birth year had the major effect on data cut-off, especially for FRA and GBR
- Trait-Breed-Sex biological ranges may be needed to have better edits
- With single-breed models, heterosis and %purity columns may not be needed.

N. common bulls, CHA:

	CZE	DNK	FIN	FRA	IRL	SWE
CZE	804					
DNK	47	1,038				
FIN	40	31	834			
FRA						
IRL	34	57	17		2,718	
SWE	35	47	91		25	3,983

N. progeny from common bulls in the 2 countries, CHA:

	CZE	DNK	FIN	FRA	IRL	SWE
CZE	24371					
DNK	899431	15564				
FIN	2156646	345513	17185			
FRA						
IRL	4246416	704660	1692906		20216	
SWE	1454024	312356	626913		1105090	99691

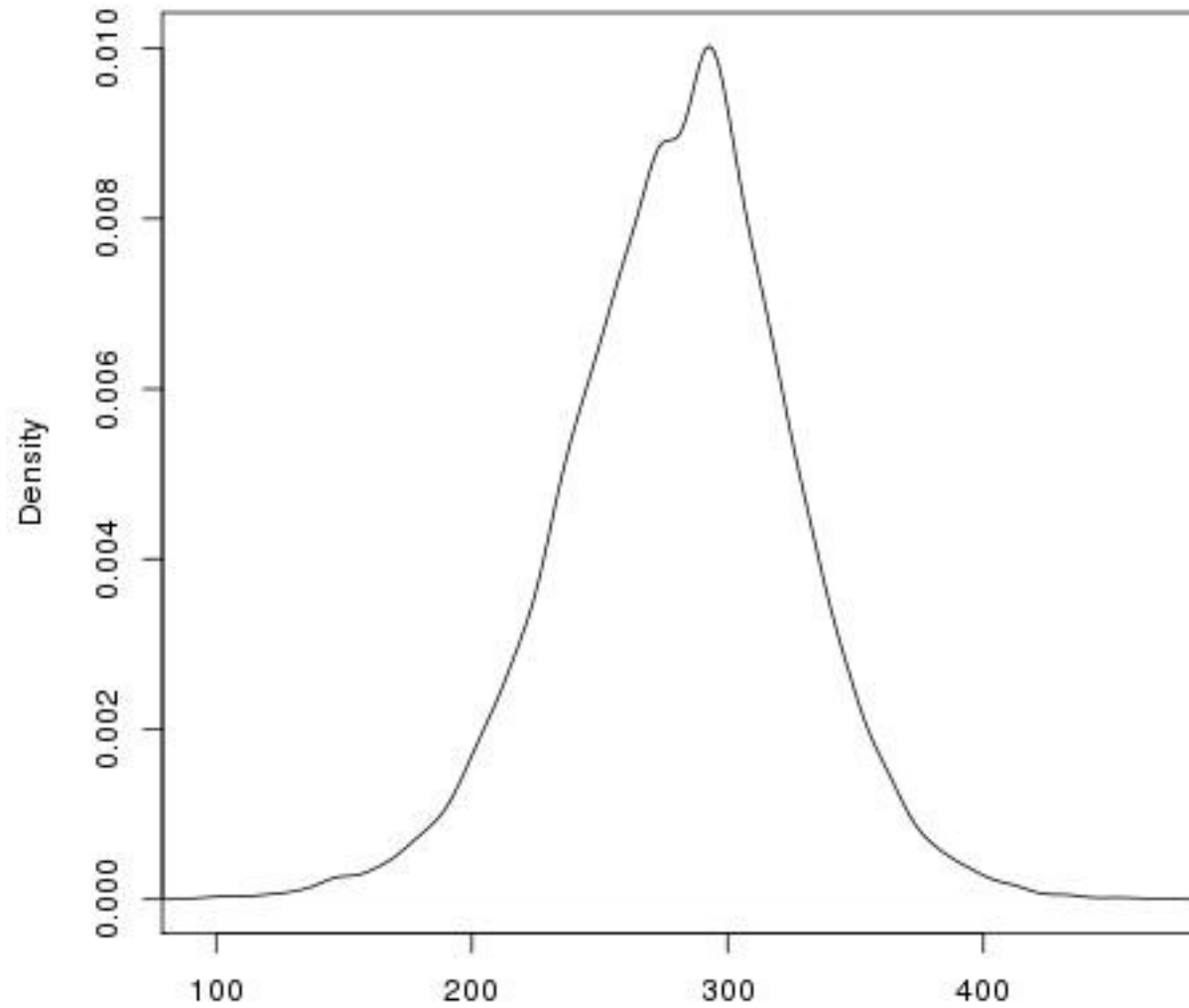
N. common bulls, LIM:

	CZE	DNK	ESP	FIN	FRA	GBR	IRL	SWE
CZE	281							
DNK	38	2,274						
ESP	43	67	1,190					
FIN	26	56	31	618				
FRA								
GBR	40	75	99	47		4,852		
IRL	33	61	78	29		179	2,099	
SWE	22	66	32	37		30	21	857

N. progeny from common bulls in the 2 countries, LIM:

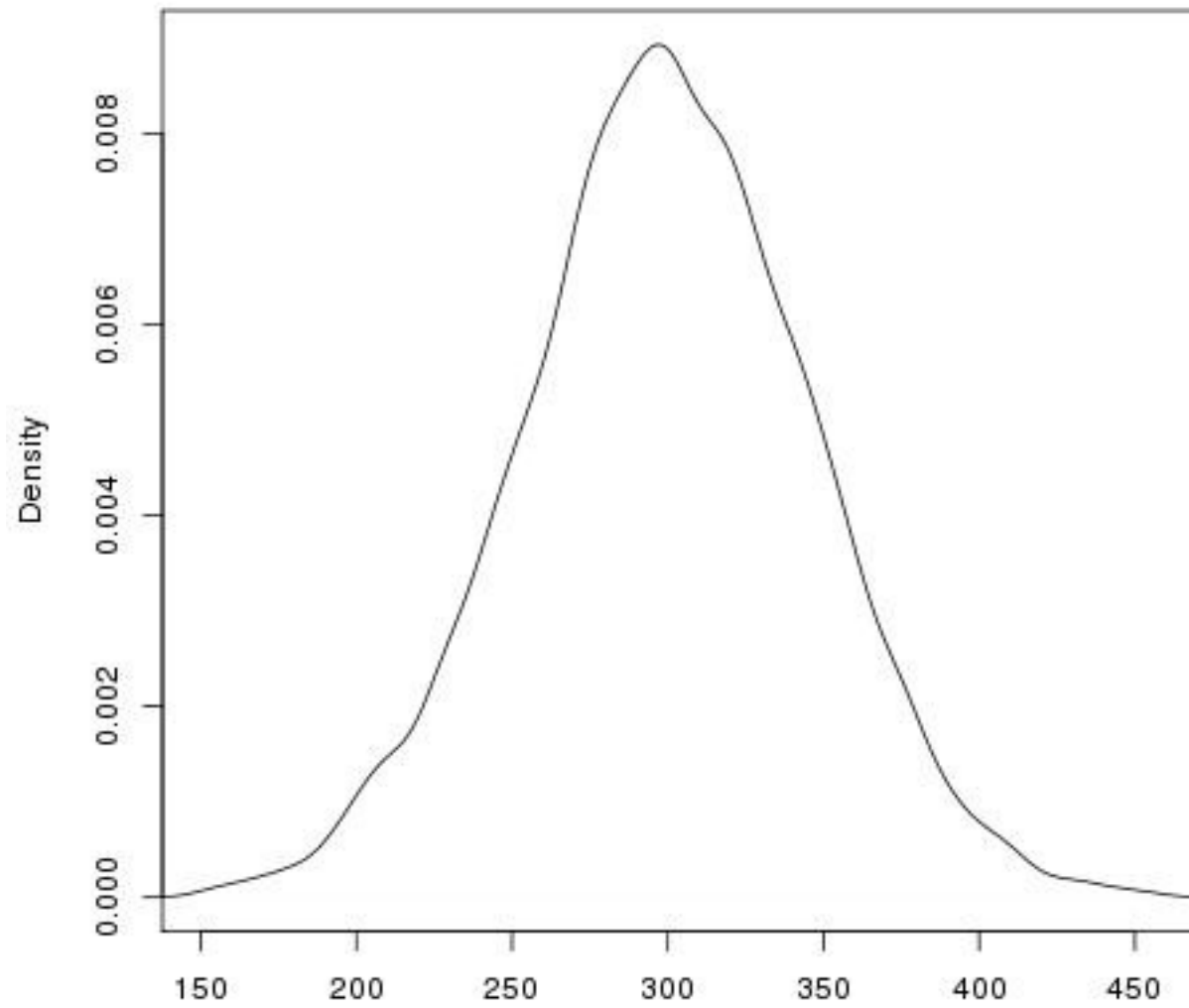
	CZE	DNK	ESP	FIN	FRA	GBR	IRL	SWE
CZE	5796							
DNK	617439	40737						
ESP	336700	478476	33259					
FIN	244744	370212	146216	13036				
FRA								
GBR	2410211	2659949	1511076	1066358		112845		
IRL	755931	921713	672309	332970		4222641	18806	
SWE	425416	562921	246708	18804		1836101	539237	18201

Kernel Density Plot for CZE-CHA-AWW



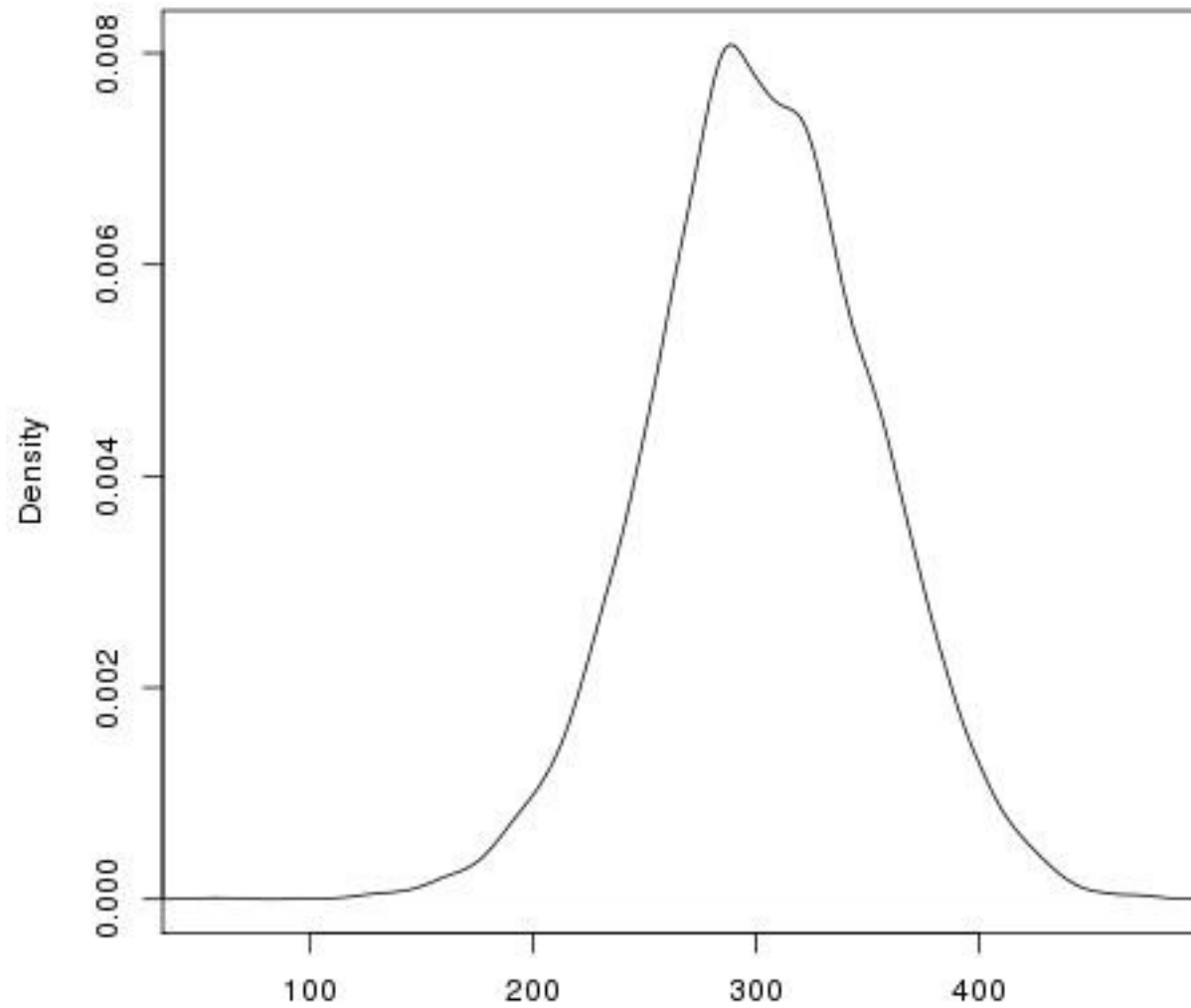
N = 16216 Bandwidth = 5.508

Kernel Density Plot for DNK-CHA-AWW



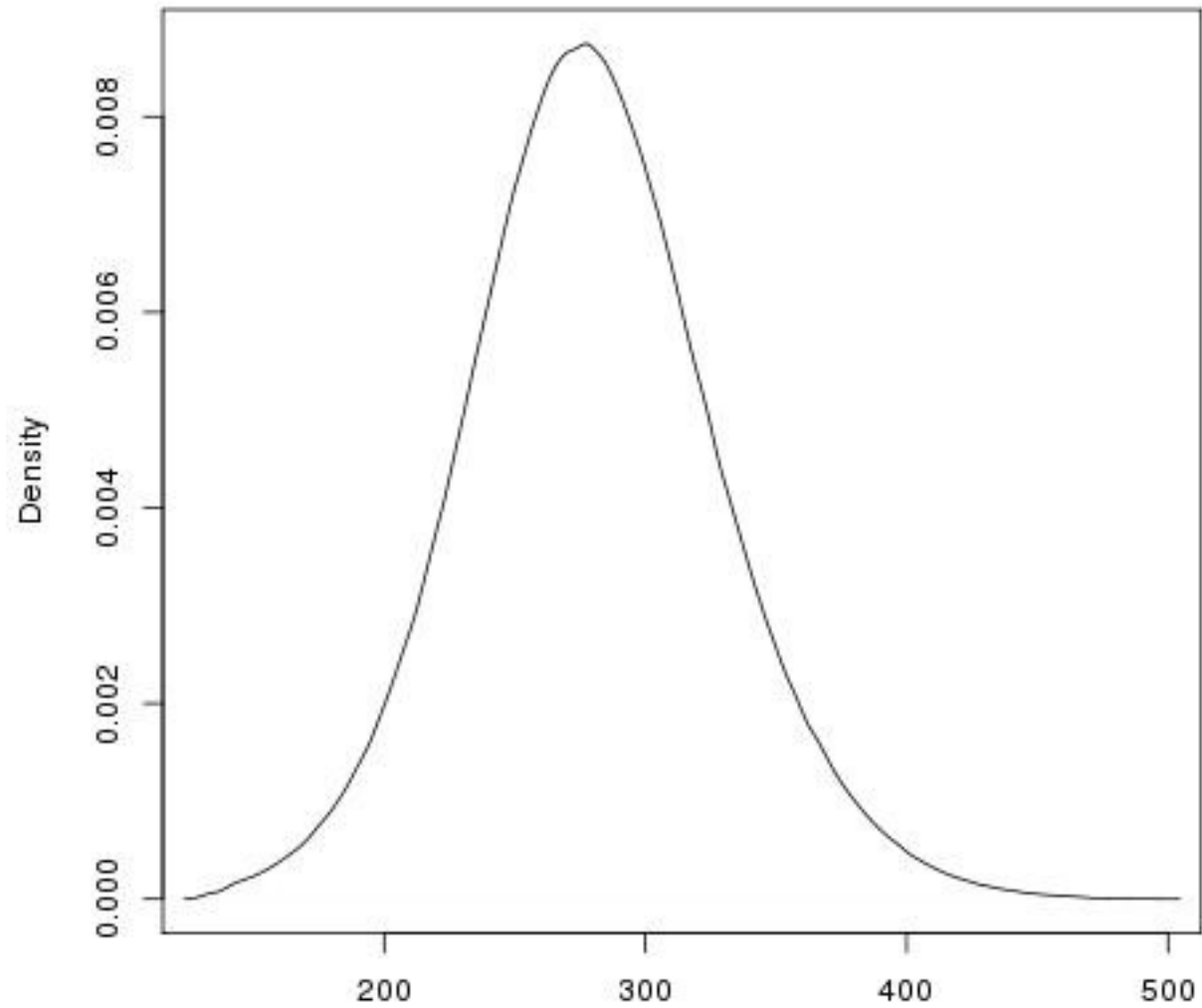
N = 10098 Bandwidth = 6.374

Kernel Density Plot for FIN-CHA-AWW



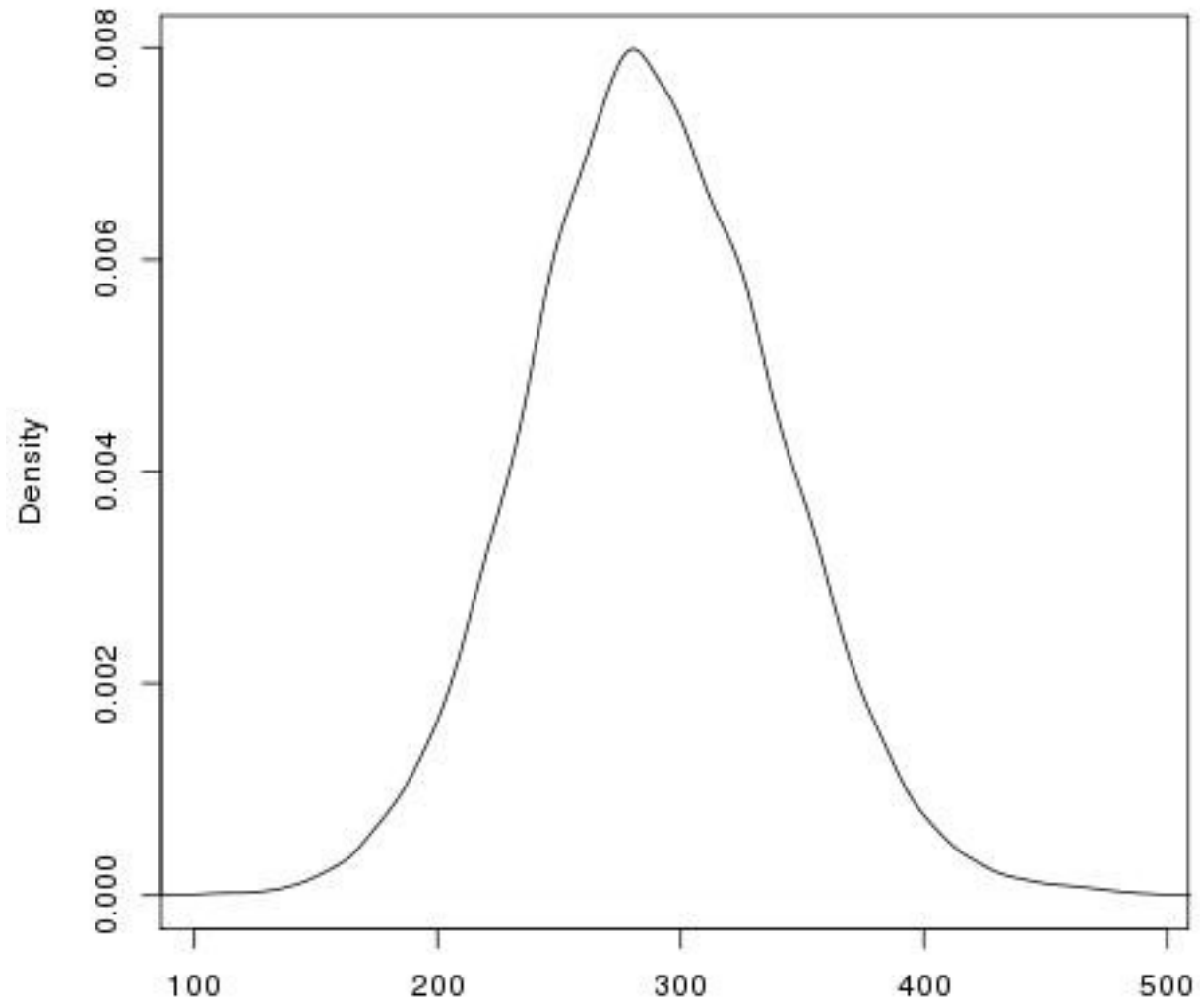
N = 13680 Bandwidth = 6.699

Kernel Density Plot for FRA-CHA-AWW



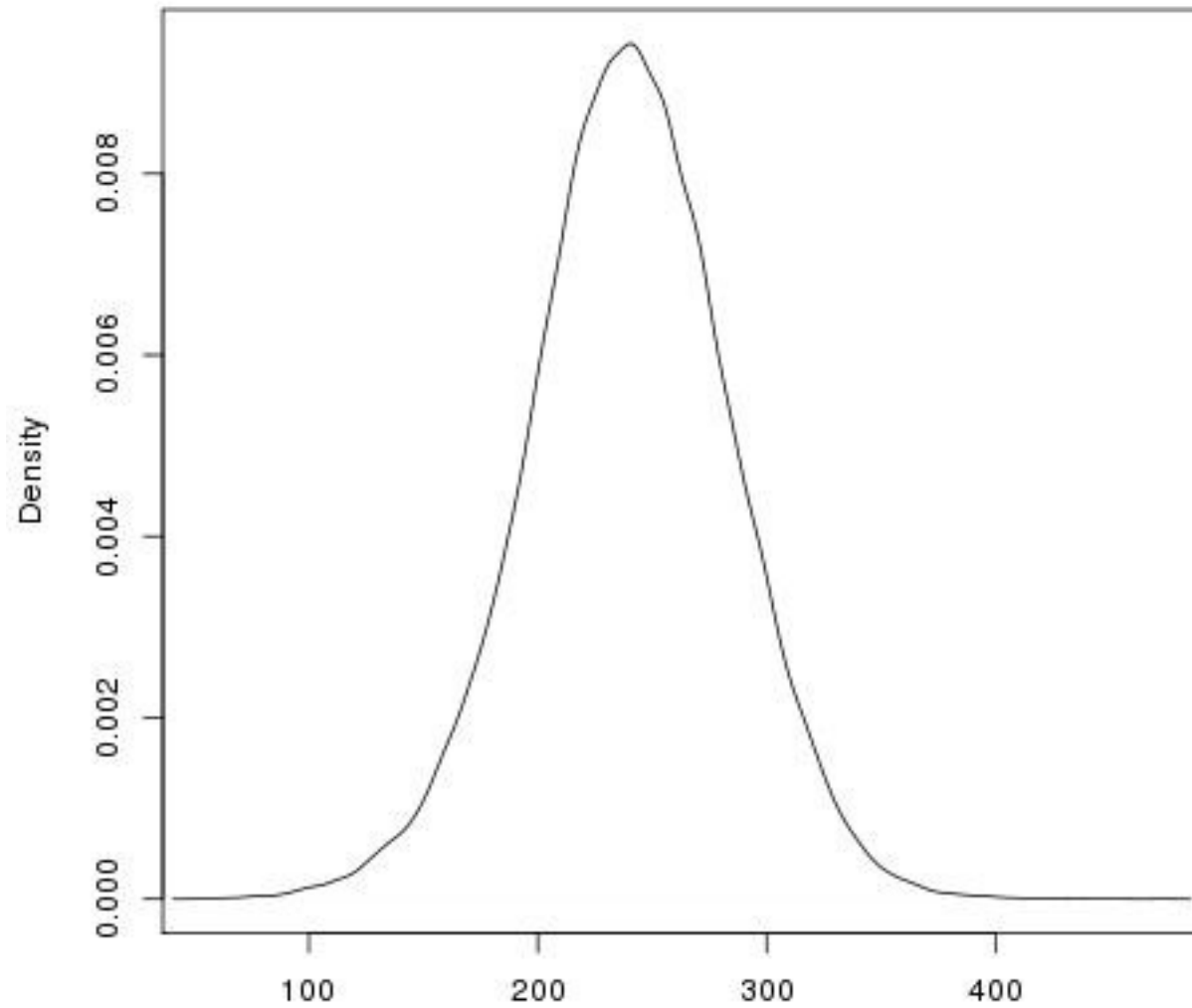
N = 2780326 Bandwidth = 2.141

Kernel Density Plot for IRL-CHA-AWW



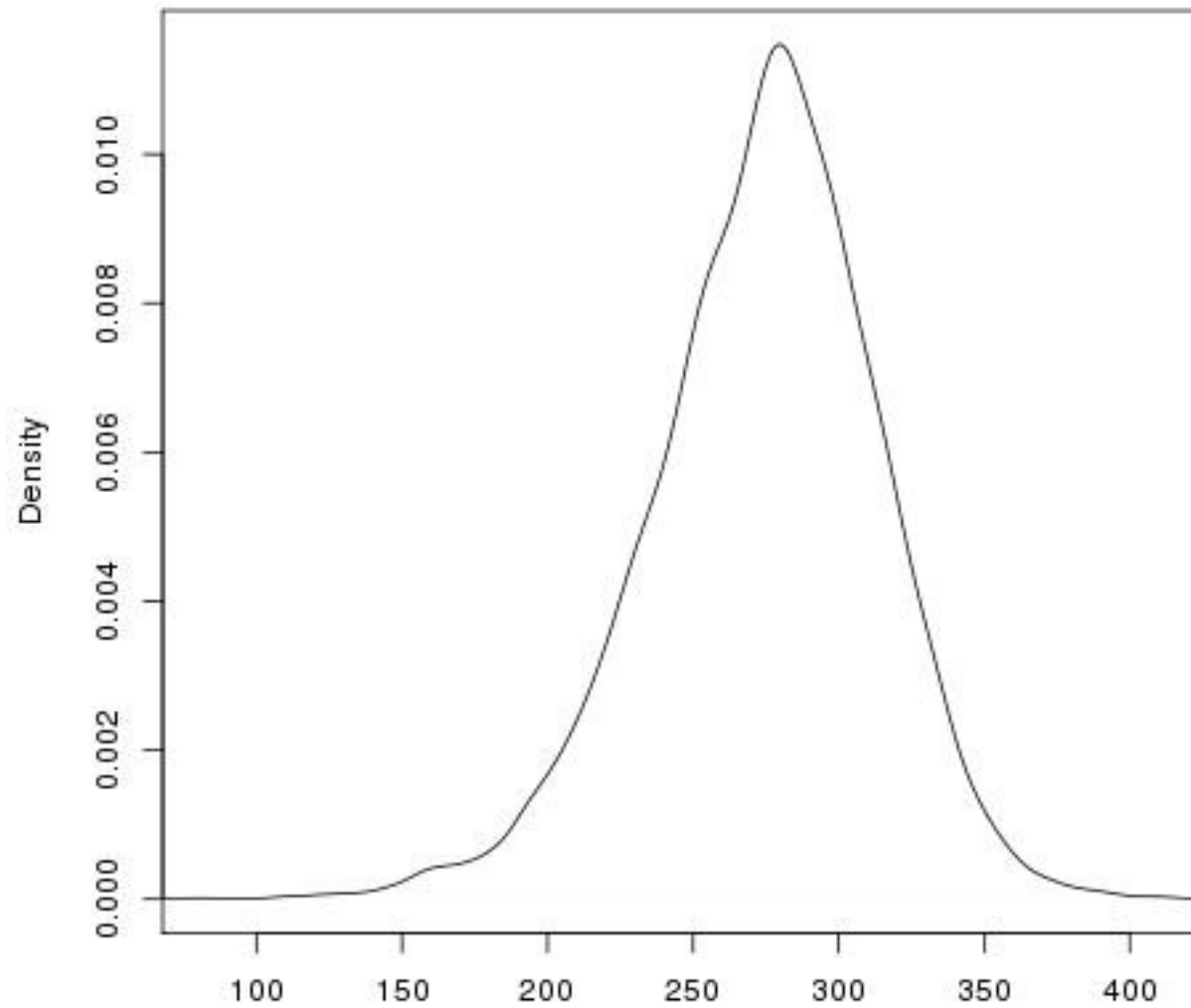
N = 13961 Bandwidth = 6.849

Kernel Density Plot for SWE-CHA-AWW



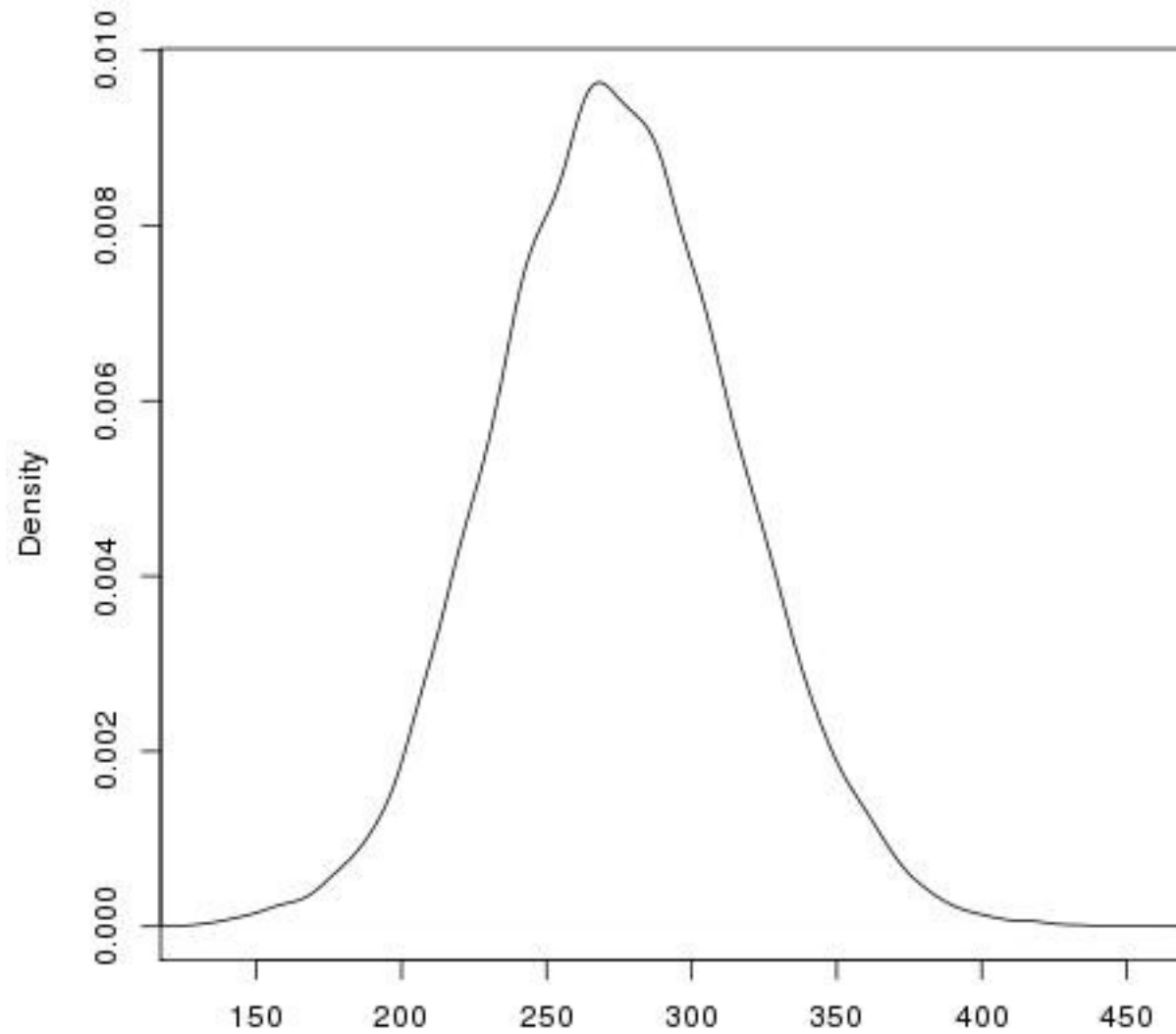
N = 79061 Bandwidth = 4.013

Kernel Density Plot for CZE-LIM-AWW



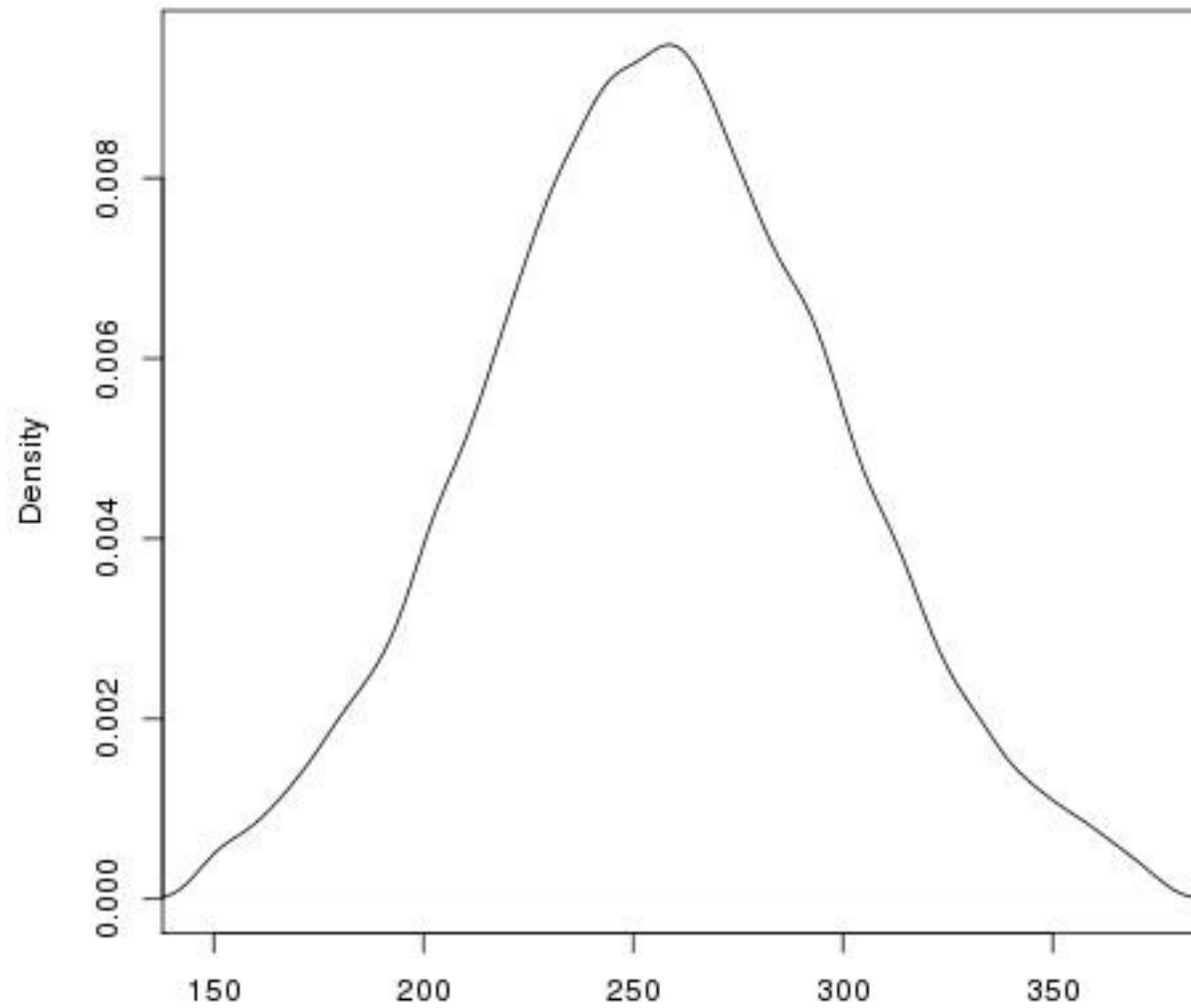
N = 4321 Bandwidth = 6.169

Kernel Density Plot for DNK-LIM-AWW



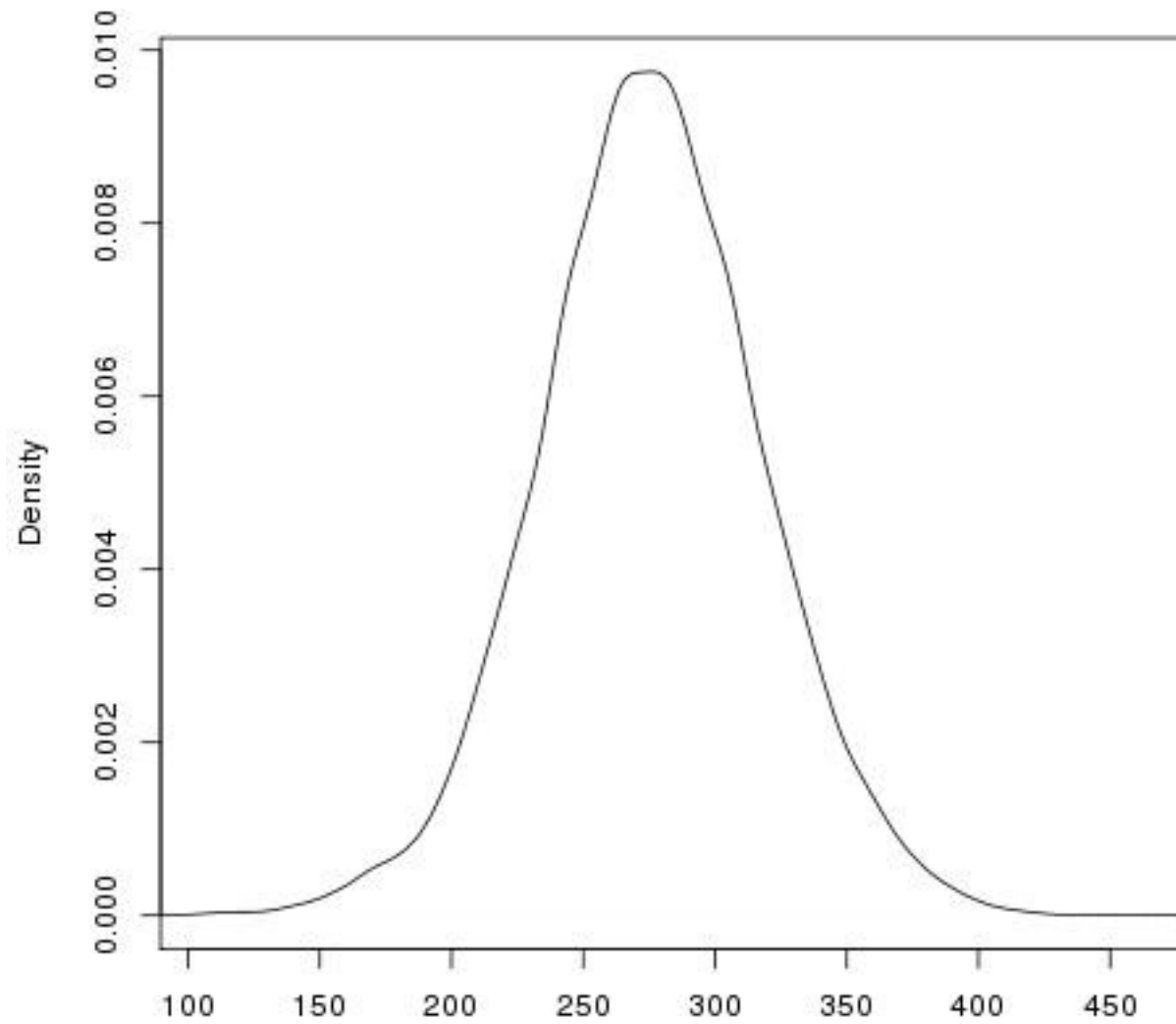
N = 30061 Bandwidth = 4.783

Kernel Density Plot for ESP-LIM-AWW



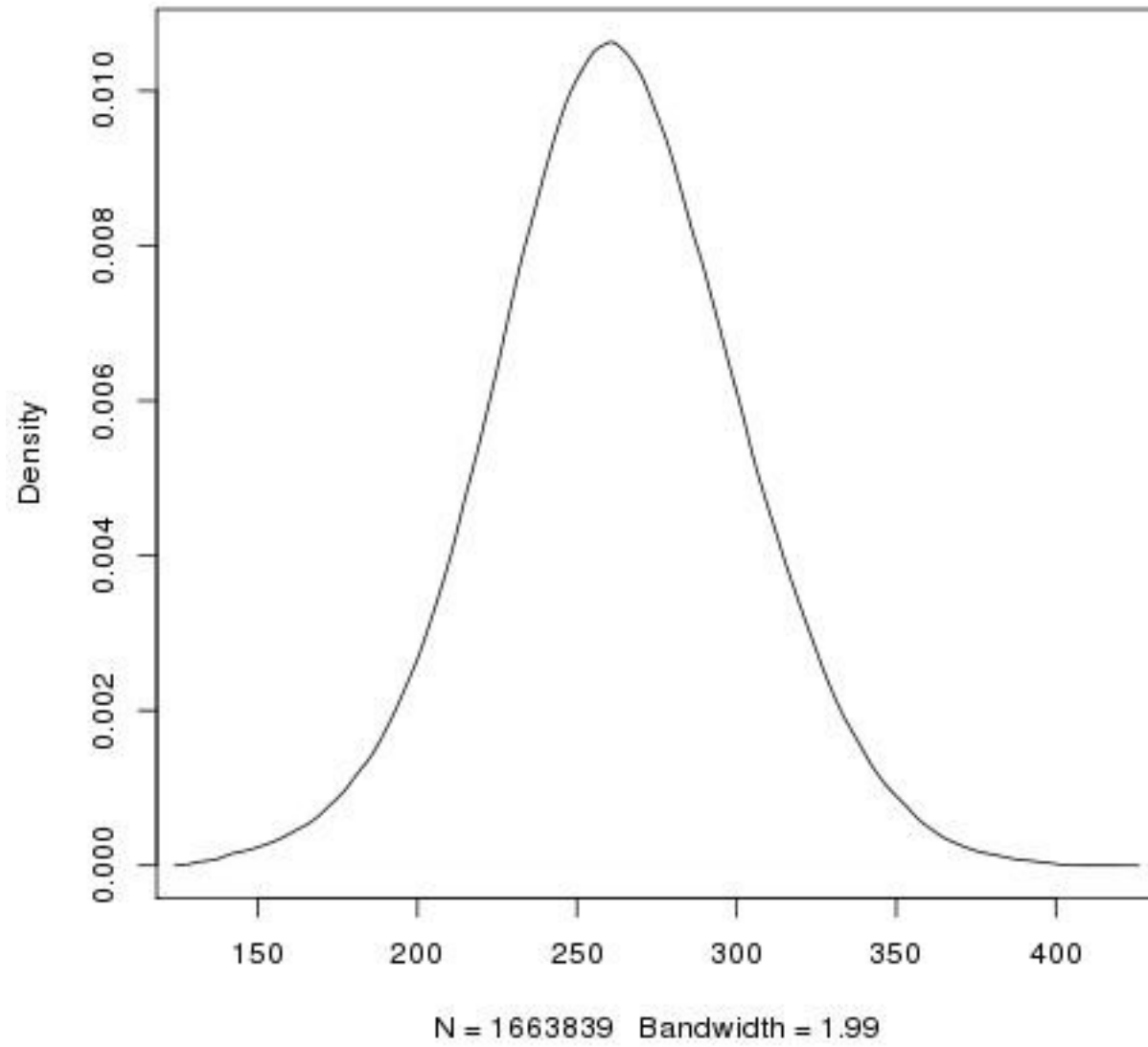
N = 32387 Bandwidth = 4.801

Kernel Density Plot for FIN-LIM-AWW

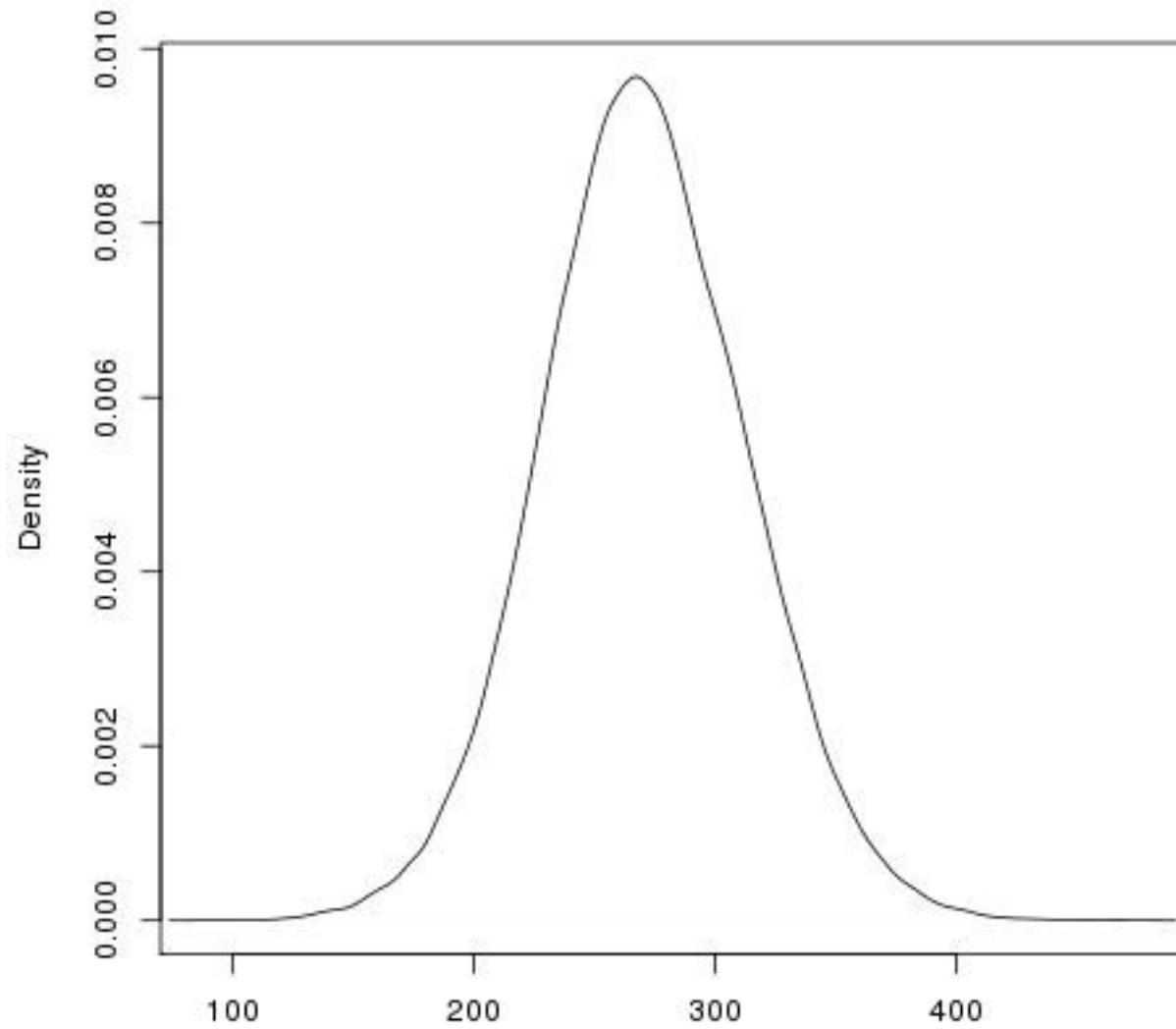


N = 11136 Bandwidth = 5.73

Kernel Density Plot for FRA-LIM-AWW

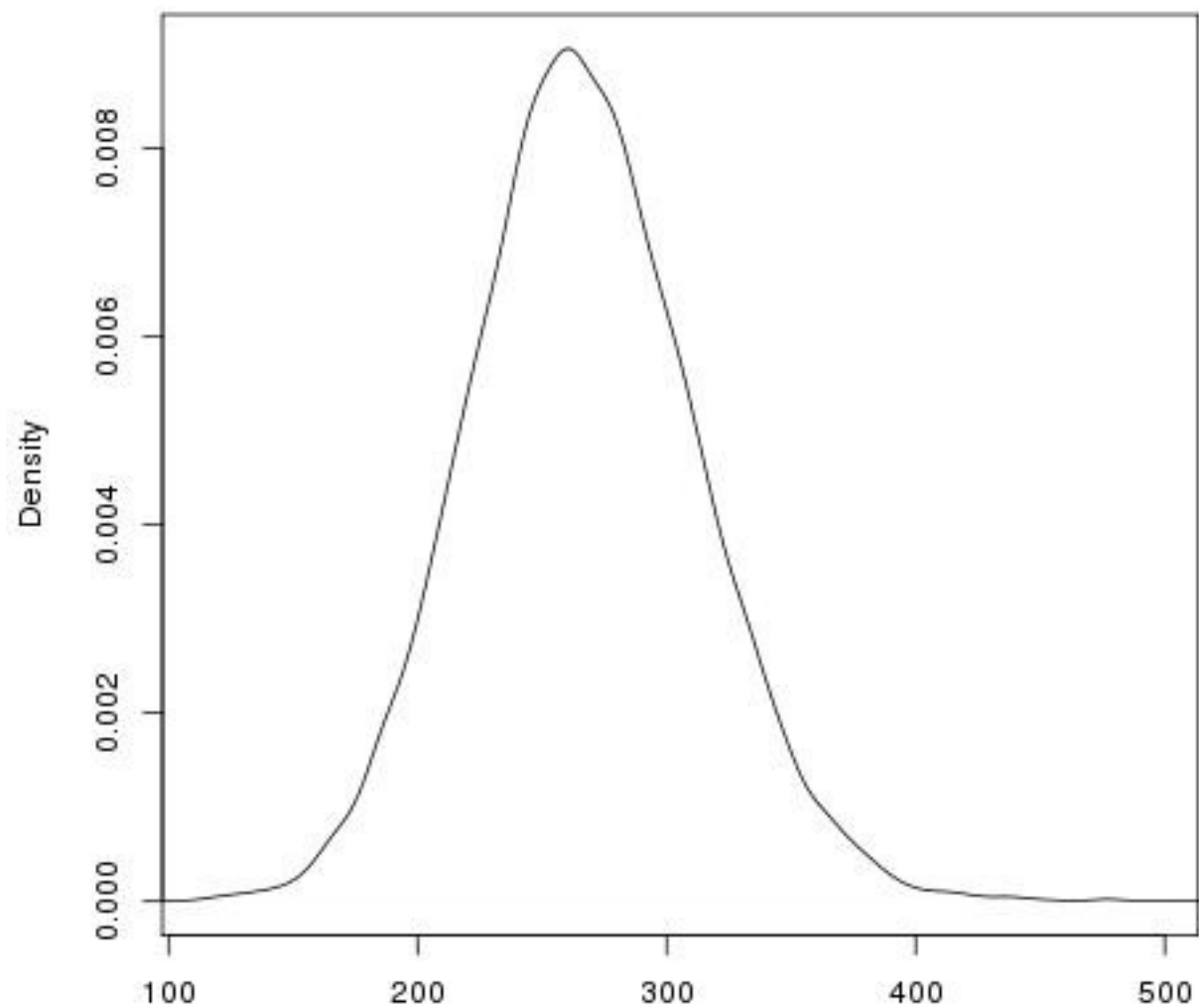


Kernel Density Plot for GBR-LIM-AWW



N = 81493 Bandwidth = 3.988

Kernel Density Plot for IRL-LIM-AWW



N = 13869 Bandwidth = 5.982

Kernel Density Plot for SWE-LIM-AWW

