log GWAS: Name FB 7/12 → 3/6

1. Formatting:

0.75/1

all margins 2.5cm	informative title
12 pt size	name on all pages
no raw R code or output	all pages numbered
max **10** pages	no blurry plots (**NOT png**)

2. Introduction/Background:

0.75/1

brief background and statement of scientific question (clear)

all variables defined

3. EDA / PCA: meanings of Tables 1 & 2 unclear

1.25/2

- vague explanations - 'problems' with sample?

relevant histograms / exploratory plots (NO BOXPLOTS)

Clearly ***EXPLAIN RELATION*** between PCs and population stratification

+ Do you do PC on Cov or corr matrix?

plot pc2 (y-axis) vs pc1 (x-axis)

4. Pre-processing / QC steps: CLEARLY EXPLAIN don't just state/conclude

0.75/2

SNP QC: criteria and reasons

sample QC: criteria and reasons

Hardy-Weinberg equilibrium: DEFINE and say how it relates to quality (not just random mating) incorrect conclusion + not clear explanation

measure of LD and how it is used

5. Association / post-association analysis: + give PRIMARY references + Devlin + Roeder

1/2

write out model ***mathematically*** (for a given SNP): CLEARLY define all variables and domains - Model MUST RELATE TO SNP + age + sex-adjusted

describe association analysis in words and mathematically

Don't make it square

- Manhattan plot and explanation horizontal lines?

lambda analysis (***including*** mathematical definition of lambda and **SQUARE** QQ plots) incomplete

(caption: not 'squared')

+ explain thresholds

4.5/8

0.5/1 make a table

6. Identify significant SNPs (include all relevant information)

0.75/1 7. Plots:

label size (not too small) informative captions

Figure 7 is PC2 vs PC1

placement put PC + scree side by side explanations

0.5/1 8. Conclusions

seems over-interpreted in parts

1. recap analysis 2. state and interpret main findings

0.75/1 9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: - Defn of r^2: what is 'genotype' - it's NOT numeric

2.5/4

GWAS: Name LB 7/12 → 3.5/6

1. Formatting:

all margins 2.5cm	informative title
12 pt size	name on all pages
no raw R code or output	all pages numbered
max **10** pages	no blurry plots (**NOT png**)

0.75/1

2. Introduction/Background:

1/1 brief background and statement of scientific question

all variables defined

3. EDA / PCA:

1/2 relevant histograms / exploratory plots (NO BOXPLOTS)

Meanings of Tables 1 + 2 not clear

- other relevant plots?

EXPLAIN RELATION between PCs and population stratification

+ Do you do PC on corr or cov matrix?

plot pc2 (y-axis) vs pc1 (x-axis) - square

+ explain reasons for plots

1.25/2 4. Pre-processing / QC steps: CLEARLY EXPLAIN

SNP QC: criteria and reasons - explain (don't just conclude)

sample QC: criteria and reasons

Hardy-Weinberg equilibrium: DEFINE and say how it relates to quality

explain

measure of LD and how it is used

5. Association / post-association analysis: + give PRIMARY references

1/2 write out model ***mathematically*** (for a given SNP): CLEARLY define all variables and domains - Model MUST RELATE TO SNP

= 0/1/2 for SNP

+ age + sex - adjusted

describe association analysis in words and mathematically

Manhattan plot and explanation - incomplete - horizontal lines?

incomplete - what is denom?

lambda analysis (***including*** mathematical definition of lambda and **SQUARE** QQ plots)

+ explain thresholds

5/8

0.5/1 incomplete? SE

6. Identify significant SNPs (include all relevant information)

7. Plots:

0.75/1 label size (not too small) informative captions

placement explanations

0.25/1 8. Conclusions

stat sig not 'relevant'

1. recap analysis 2. state and interpret main findings — vague

0.5/1 9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

Breheny

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: (a few apparently unattributed quotes)

2/4

log GWAS: Name _____ CD 5.75/12 → 2.875/6

1. Formatting:

all margins 2.5cm	informative title
12 pt size	name on all pages
no raw R code or output	all pages numbered
max **10** pages	no blurry plots (**NOT png**)

0.5/1 no R commands - too many digits

2. Introduction/Background:

1/1 brief background and statement of scientific question

all variables defined Don't need plink files details

3. EDA / PCA: Don't need tables 4/9/10

relevant histograms / exploratory plots (NO BOXPLOTS) + Explain clearly reasons for plots. other relevant plots

1/2 ***EXPLAIN RELATION*** between PCs and population stratification (clearly)

plot pc2 (y-axis) vs pc1 (x-axis) + do you do PC with corr or cov matrix?

(*) - scree plot; rotate y-axis labels (see ?par

0.5/2 4. Pre-processing / QC steps: CLEARLY EXPLAIN

SNP QC: criteria and reasons homogene (very superficial) not 'ensure' and look at las parameter) + explain

sample QC: criteria and reasons all gts

Hardy-Weinberg equilibrium: DEFINE and say how it relates to quality (explain

measure of LD and how it is used

validation?

0.5/2 5. Association / post-association analysis: + give PRIMARY references

write out model ***mathematically*** (for a given SNP): CLEARLY define all variables and domains - Model MUST RELATE TO SNP

describe association analysis in words and mathematically post-FDR not Defined

Manhattan plot and explanation - horizontal lines + explain thresholds

lambda analysis (***including*** mathematical definition of lambda and **SQUARE** QQ plots) + interpret IGC (and cite)

(*) + scree plot and next plot redundant

3.5/8

0.5/1 6. Identify significant SNPs (include all relevant information) Effect sizes and SES

0.75/1 7. Plots:

label size (not too small) Q-Q (+number all figures

informative captions

placement Manhattan explanations

0.5/1 8. Conclusions

1. recap analysis (First, before any result) 2. state and interpret main findings (likely?)

0.5/1 9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory very incomplete good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: _____

_____

_____

_____

_____

_____

2.25/4

GWAS: Name _____

1. Formatting:

all margins 2.5cm	informative title
12 pt size	name on all pages
no raw R code or output	all pages numbered
max **10** pages	no blurry plots (**NOT png**)

2. Introduction/Background:

brief background and statement of scientific question

all variables defined

3. EDA / PCA:

relevant histograms / exploratory plots (NO BOXPLOTS)

EXPLAIN RELATION between PCs and population stratification

plot pc2 (y-axis) vs pc1 (x-axis)

4. Pre-processing / QC steps: CLEARLY EXPLAIN

SNP QC: criteria and reasons

sample QC: criteria and reasons

Hardy-Weinberg equilibrium: DEFINE and say how it relates to quality

measure of LD and how it is used

5. Association / post-association analysis: + give PRIMARY references

write out model ***mathematically*** (for a given SNP): CLEARLY define all variables and domains - Model MUST RELATE TO SNP

describe association analysis in words and mathematically

Manhattan plot and explanation

lambda analysis (***including*** mathematical definition of lambda and **SQUARE** QQ plots)

0.5/1 in table

6. Identify significant SNPs (include all relevant information)

0.5/1 7. Plots:

label size (not too small) informative captions

placement QQ in body of report + explanations

0.25/1 8. Conclusions

not 'confirmed'

1. recap analysis 2. state and interpret main findings

0.5/1 9. Overall presentation (clarity of explanations, appropriate citations / references) : spell check

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: _____

_____

_____

_____

_____

_____

1.75/4

log GWAS: Name _____ EF 6.75/12 → 3.875/6

1. Formatting:

all margins 2.5cm	informative title
12 pt size	name on all pages
no raw R code or output	all pages numbered
max **10** pages	no blurry plots (**NOT png**)

0.75/1

2. Introduction/Background:

brief background and statement of scientific question

PennCATH is the study (PennCATH cohort is the cohort not the study) of your report, specifically

all variables defined

0.75/1

3. EDA / PCA: + explanations - why do you make these plots?

relevant histograms / exploratory plots (NO BOXPLOTS)

- clearly explain reasons for plots

EXPLAIN RELATION between PCs and population stratification

clearly

+ Do you do PC on Corr or Cov matrix?

plot pc2 (y-axis) vs pc1 (x-axis) - square

1/2

4. Pre-processing / QC steps: CLEARLY EXPLAIN

SNP QC: criteria and reasons - be explicit in explanations not 'ensure'

sample QC: criteria and reasons + vague statement

Hardy-Weinberg equilibrium: DEFINE and say how it relates to quality - explain, don't just state

measure of LD and how it is used

not just random mating - all genotypes

+ Zeylin + Reeder

0.75/2

5. Association / post-association analysis: + give PRIMARY references

write out model ***mathematically*** (for a given SNP): CLEARLY define all variables and domains - Model MUST RELATE TO SNP

snp + sex numerically

describe association analysis in words and mathematically

Manhattan plot and explanation incomplete - horizontal lines + explain thresholds

lambda analysis (***including*** mathematical definition of lambda and **SQUARE** QQ plots)

incomplete - what is denom?

1/2

9.25/8

0.5/1 + put in table

6. Identify significant SNPs (include all relevant information)

7. Plots:

0.75/1 label size (not too small) informative captions

placement explanations

0.75/1 8. Conclusions highly sig genetic signal

1. recap analysis 2. state and interpret main findings

(be more complete)

0.5/1 9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references Brehemy

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: _____

_____

_____

_____

_____

_____

2.5/4

log

GWAS: Name LF 6/12 → 3/6

1. Formatting:

(Strange section numbering)

all margins 2.5cm — informative title

12 pt size — name on all pages

no raw R code or output — all pages numbered

max **10** pages — no blurry plots (**NOT png**)

0.5/1

2. Introduction/Background:

brief background and statement of scientific question

all variables defined

1/1

3. EDA / PCA:

- clearly explain reasons for plots
Don't need table in 0.2

relevant histograms / exploratory plots (NO BOXPLOTS)

+ explain PC methodology

EXPLAIN RELATION between PCs and population stratification

+ Do you do PC on Corr or Cov matrix?

plot pc2 (y-axis) vs pc1 (x-axis)

0.75/2

4. Pre-processing / QC steps: CLEARLY EXPLAIN

- Table Contents not explained

SNP QC: criteria and reasons

Very superficial

sample QC: criteria and reasons

(not only random mating)

Hardy-Weinberg equilibrium: DEFINE and say how it relates to quality

measure of LD and how it is used

not defined

0.25/2

5. Association / post-association analysis: + give PRIMARY references

- explain + Devlin + Roeder

write out model ***mathematically*** (for a given SNP): CLEARLY define all variables and domains - Model MUST RELATE TO SNP

+ age + sex adjusted

describe association analysis in words and mathematically

(what is 'ref', 'alt'?)

Manhattan plot and explanation

incomplete - explain thresholds + horizontal lines

lambda analysis (***including*** mathematical definition of lambda and SQUARE QQ plots)

incomplete - what is denom?

0.75/2

3.25/8

0.5/1 put in table

6. Identify significant SNPs (include all relevant information)

7. Plots:

0.75/1 + number all figures

label size (not too small) informative captions

placement explanations

8. Conclusions

1/1

1. recap analysis 2. state and interpret main findings

9. Overall presentation (clarity of explanations, appropriate citations / references) :

0.5/1

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: - number all figures and tables

* very superficial report - you need to explain reasons, not just say 'we did this' and conclude

2.75/4

Reed GWAS: Name JL 6.5/12 → 3.25/6

1. Formatting:

0.75/1

all margins 2.5cm — informative title

12 pt size — name on all pages

no raw R code or output (no R commands) — all pages numbered

max **10** pages — no blurry plots (**NOT png**) – Too many digits

0.5/1 2. Introduction/Background:

brief background and statement of scientific question — you talk about CAD but you specifically analyzed (trans)HDL

all variables defined — Don't need plink files details

Spell out words

3. EDA / PCA: Don't need Tables 1/2/3

1/2 relevant histograms / exploratory plots (NO BOXPLOTS) — A1 & A2 in same plot?

EXPLAIN RELATION between PCs and population stratification — do you do PC with corr or cov matrix?

plot pc2 (y-axis) vs pc1 (x-axis) — other relevant plots?

4. Pre-processing / QC steps: CLEARLY EXPLAIN — not 'ensure'

0.5/2 SNP QC: criteria and reasons — superficial explanations — how are the measures relate to reliability / quality

sample QC: criteria and reasons — all gts

Hardy-Weinberg equilibrium: DEFINE and say how it relates to quality

measure of LD and how it is used — not explained

5. Association / post-association analysis: + give PRIMARY references — Devlin + Roeder

1/2 write out model ***mathematically*** (for a given SNP): CLEARLY define all variables and domains - Model MUST RELATE TO SNP — what is Sex (numerically)?

describe association analysis in words and mathematically — Bonferroni?

Manhattan plot and explanation — ? QC? — (clearly) - where do the p-values come from?

lambda analysis (***including*** mathematical definition of lambda and **SQUARE** QQ plots)

incomplete - where do the thresholds come from?

3.75/8

6. Identify significant SNPs (include all relevant information)

7. Plots:

label size (not too small) informative captions

placement explanations

8. Conclusions

1. recap analysis 2. state and interpret main findings

9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other:

GWAS: Name _____ JL-F 3.25/12 → 1.625/6

1. Formatting:

NO AI

all margins 2.5cm	informative title
12 pt size	name on all pages
no raw R code or output	all pages numbered
max **10** pages	no blurry plots (**NOT png**)

0/1

2. Introduction/Background: not 'complex', continuous (the diseases you mention are considered complex)

0.5/1

brief background and statement of scientific question — imprecise + incomplete

all variables defined

3. EDA / PCA: Don't need R details or plink file details

0.75/2

Don't need Figure 4

relevant histograms / exploratory plots (NO BOXPLOTS)

EXPLAIN RELATION between PCs and population stratification — don't just state

plot pc2 (y-axis) vs pc1 (x-axis)

4. Pre-processing / QC steps: CLEARLY EXPLAIN — Superficial

0.5/2

SNP QC: criteria and reasons — explain explicitly

sample QC: criteria and reasons

Hardy-Weinberg equilibrium: DEFINE and say how it relates to quality — explain

measure of LD and how it is used — incomplete + all gts

5. Association / post-association analysis: + give PRIMARY references

0.75/2

write out model ***mathematically*** (for a given SNP): CLEARLY define all variables and domains - Model MUST RELATE TO SNP — you don't correct for age + sex

describe association analysis in words and mathematically → what do gt=0/1/2 represent?

Manhattan plot and explanation — no solid line plot'n dotted — explain thresholds

lambda analysis (***including*** mathematical definition of lambda and **SQUARE** QQ plots) — incomplete

what hyp test / where do p-values come from?

2.5/8

0/1 Put results in a table

6. Identify significant SNPs (include all relevant information)

0.75/1 7. Plots: Some plots too big PC + Scree side by side

label size (not too small) informative captions

placement explanations

0/1 8. Conclusions not done?

1. recap analysis 2. state and interpret main findings

0/1 9. Overall presentation (clarity of explanations, appropriate citations / references) : (D) + rather incomplete + NO AI

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

(*) D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: (*) Substantial Copying from Brebeny tutorial

0.75/4

Real GWAS: Name _____ WL 7.25/12 → 3.625/6

1. Formatting:

all margins 2.5cm	informative title
12 pt size	name on all pages
no raw R code or output	all pages numbered
max **10** pages	no blurry plots (**NOT png**)

1/1

2. Introduction/Background: Define GWA before using abbreviation

1/1

brief background and statement of scientific question

all variables defined (M M / F F) not M F M F

3. EDA / PCA:

1/2 + other relevant plots?

relevant histograms / exploratory plots (NO BOXPLOTS)

EXPLAIN RELATION between PCs and population stratification

+ do you do PC with cov or cor matrix? (Square)

plot pc2 (y-axis) vs pc1 (x-axis)

LD and r^2 not explained

1/2 4. Pre-processing / QC steps: CLEARLY EXPLAIN don't just cite refs

SNP QC: criteria and reasons

sample QC: criteria and reasons all gts

Hardy-Weinberg equilibrium: DEFINE and say how it relates to quality

measure of LD and how it is used measure not defined

5. Association / post-association analysis: + give PRIMARY references

(*) what is rank-normalized TG? Why z-trans age?

write out model ***mathematically*** (for a given SNP): CLEARLY define all variables and domains - Model MUST RELATE TO SNP

0.75/2 describe association analysis in words and mathematically - What hypothesis test?

Manhattan plot and explanation incomplete

lambda analysis (***including*** mathematical definition of lambda and **SQUARE** QQ plots) incomplete - what is denom?

(*) Clearly define what GT means

4.75/8

0.5/1 Make a table

6. Identify significant SNPs (include all relevant information)

0.75/1 7. Plots:

label size (not too small) informative captions what are the horizontal lines in the Manhattan plots?

placement (put PC + Scree explanations side by side)

8. Conclusions not 'guarantee rigor'

0.75/1 1. recap analysis 2. state and interpret main findings how underscore? somewhat vague and unclear

0.5/1 9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references Reed et al.

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: _____

_____

_____

_____

_____

_____

2.5/4

GWAS: Name JN 7.75/12 → 3.875/6

1. Formatting:

all margins 2.5cm	informative title
12 pt size	name on all pages
no raw R code or output	all pages numbered
max **10** pages	no blurry plots (**NOT png**)

2. Introduction/Background: not clear not PEER-reviewed

brief background and statement of scientific question - clearly

all variables defined

3. EDA / PCA:

- explain reasons for these plots

relevant histograms / exploratory plots (NO BOXPLOTS)

clearly

EXPLAIN RELATION between PCs and population stratification

+ Do you do PC on corr or cov matrix?

plot pc2 (y-axis) vs pc1 (x-axis)

4. Pre-processing / QC steps: CLEARLY EXPLAIN not 'ensure'

somewhat superficial

SNP QC: criteria and reasons - explain why

sample QC: criteria and reasons conditions?

Hardy-Weinberg equilibrium: DEFINE and say how it relates to quality explain, don't just conclude

measure of LD and how it is used

+ Devlin + Roeder

5. Association / post-association analysis: + give PRIMARY references

write out model ***mathematically*** (for a given SNP): CLEARLY define all variables and domains - Model MUST RELATE TO SNP

describe association analysis in words and mathematically (+ alt hyp)

Manhattan plot and explanation - explain thresholds

lambda analysis (***including*** mathematical definition of lambda and **SQUARE** QQ plots) incomplete - what is denom?

5/8

0.5/1

6. Identify significant SNPs (include all relevant information) est β's + SEs + p

7. Plots:

0.75/1

label size (not too small) informative captions

placement PC + Scree side by side explanations

0.5/1

8. Conclusions

1. recap analysis 2. state and interpret main findings not 'confirmed' explain

1/1

9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments: Breheny

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: _____

2.75/4

log GWAS: Name _____ NN 6.25/12 → 3.125/6

1. Formatting:

all margins 2.5cm	informative title
12 pt size	name on all pages
no raw R code or output	all pages numbered
max **10** pages	no blurry plots (**NOT png**)

0.75/1

2. Introduction/Background:

brief background and statement of scientific question primarily?

all variables defined

1/1

3. EDA / PCA:

relevant histograms / exploratory plots (NO BOXPLOTS) clearly explain reasons for plots

EXPLAIN RELATION between PCs and population stratification clearly

plot pc2 (y-axis) vs pc1 (x-axis) + Do you do PC on Corr or Covariance matrix? Scree (side by side) (Assess, not 'verify')

1/2

4. Pre-processing / QC steps: CLEARLY EXPLAIN

SNP QC: criteria and reasons very superficial — explain

sample QC: criteria and reasons

Hardy-Weinberg equilibrium: DEFINE and say how it relates to quality not only random mating all g's explain - don't just conclude

measure of LD and how it is used — explain + $r^2 = ?$ + Taylin + Roeder Define numerically

0.5/2

5. Association / post-association analysis: + give PRIMARY references

write out model ***mathematically*** (for a given SNP): CLEARLY define all variables and domains - Model MUST RELATE TO SNP [+ age + sex + snp]

describe association analysis in words and mathematically Define logit

Manhattan plot and explanation What hypotest? Where do p-values come from?

lambda analysis (***including*** mathematical definition of lambda and SQUARE QQ plots) incomplete - what is denom? incomplete horizontal lines? + explain thresholds

0.75/1

4/8

0.5/1 p's not EXACTLY =0 + don't need niter, score, log lop

6. Identify significant SNPs (include all relevant information) tail

0.75/1 7. Plots:

label size (not too small) (labels) informative captions

placement explanations

0.5/1 8. Conclusions use paragraphing

1. recap analysis 2. state and interpret main findings

0.5/1 9. Overall presentation (clarity of explanations, appropriate citations / references) : incomplete sometimes vague

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: _____

_____

_____

_____

_____

_____

2.25/4

log GWAS: Name _____ RR 7/12 → 3.5/6

1. Formatting:

all margins 2.5cm	informative title ('Associated' not 'Relevant')
12 pt size	name on all pages
no raw R code or output	all pages numbered
max **10** pages	no blurry plots (**NOT png**)

1/1

2. Introduction/Background: - EIGENSTRAT is an implementation of a method, not a method

0.75/1 brief background and statement of scientific question — unclear and imprecise

all variables defined

3. EDA / PCA: Don't really need tables 1/2/3/4

1/2 relevant histograms / exploratory plots (NO BOXPLOTS) explain reasons for these plots clearly

EXPLAIN RELATION between PCs and population stratification

+ Do you do PC on cov or cor matrix?

plot pc2 (y-axis) vs pc1 (x-axis)

4. Pre-processing / QC steps: CLEARLY EXPLAIN

4/2 SNP QC: criteria and reasons - EXPLAIN

sample QC: criteria and reasons

Hardy-Weinberg equilibrium: DEFINE and say how it relates to quality

measure of LD and how it is used + Devlin + Roeder numerically

5. Association / post-association analysis: + give PRIMARY references

1/2 write out model ***mathematically*** (for a given SNP): CLEARLY define all variables and domains - Model MUST RELATE TO SNP ~ + age + sex + SNP GT

describe association analysis in words and mathematically incomplete - explain

Manhattan plot and explanation lines + thresholds horizontal incomplete - what is

lambda analysis (***including*** mathematical definition of lambda and SQUARE QQ plots) denom?

shape, Not part of title

4.75/8

0.5/1 (all) $\hat{\beta}$, SE

6. Identify significant SNPs (include all relevant information)

7. Plots:

0.75/1 label size (not too small) informative captions

placement explanations

8. Conclusions

0.5/1 1. recap analysis estimated odds 2. state and interpret main findings

9. Overall presentation (clarity of explanations, appropriate citations / references) : spellcheck

0.5/1 poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other:

2.25/4

log GWAS: Name KS 7.75/12 → 3.875/6

1. Formatting:

all margins 2.5cm

12 pt size

no raw R code or output

max **10** pages

informative title

name on all pages

all pages numbered

no blurry plots (**NOT png**)

0.75/1

2. Introduction/Background:

brief background and statement of scientific question

all variables defined

clear

0.75/1

3. EDA / PCA:

relevant histograms / exploratory plots (NO BOXPLOTS)

+ explain reasons for these plots

EXPLAIN RELATION between PCs and population stratification

clearly

plot pc2 (y-axis) vs pc1 (x-axis)

+ Do you do PC on Corr or Cov matrix?

0.5/2

4. Pre-processing / QC steps: CLEARLY EXPLAIN

SNP QC: criteria and reasons

vague/imprecise/incorrect

not 'guarantee'

clearly and precisely explain, what you have is vague

conclusion

sample QC: criteria and reasons

Hardy-Weinberg equilibrium: DEFINE and say how it relates to quality

explain, don't just conclude

measure of LD and how it is used

explain

5. Association / post-association analysis: + give PRIMARY references

1.75/2

write out model ***mathematically*** (for a given SNP): CLEARLY define all variables and domains - Model MUST RELATE TO SNP

– what is K

describe association analysis in words and mathematically (+ alt hyp)

Manhattan plot and explanation – explain thresholds

lambda analysis (***including*** mathematical definition of lambda and **SQUARE** QQ plots)

4.5/8

0.5/1

6. Identify significant SNPs (include all relevant information)

7. Plots:

0.75/1

label size (not too small) informative captions

placement explanations

1/1

8. Conclusions

1. recap analysis 2. state and interpret main findings

1/1

9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

Breheny

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: - Don't need cover page + Contents + Appendix

3.25/4

Joy

GWAS: Name JT 7.5/12 → 3.75/6

1. Formatting:

refs don't count towards page limit

all margins 2.5cm | informative title

0.75/1 (Throughout)

12 pt size | name on all pages

no raw R code or output | all pages numbered

max **10** pages | no blurry plots (**NOT png**)

2. Introduction/Background:

0.75/1

brief background and statement of scientific question (clear)

all variables defined

3. EDA / PCA: (*) interpretation: what is a 'reasonable number'?

1/2

+ Explain purpose of these plots (clearly)

relevant histograms / exploratory plots (NO BOXPLOTS)

clearly

EXPLAIN RELATION between PCs and population stratification

+ Do you do PC on Cor or Cov matrix?

plot pc2 (y-axis) vs pc1 (x-axis)

4. Pre-processing / QC steps: CLEARLY EXPLAIN

1/2

not 'ensure'; also robustness and reliability not defined

SNP QC: criteria and reasons — explain clearly and precisely

sample QC: criteria and reasons — incomplete?

Hardy-Weinberg equilibrium: DEFINE and say how it relates to quality (explain how, don't just conclude)

measure of LD and how it is used

+ Devlin + Roeder

5. Association / post-association analysis: + give PRIMARY references

1.5/2

write out model ***mathematically*** (for a given SNP): CLEARLY define all variables and domains - Model MUST RELATE TO SNP — what is SNP 'genotype'

describe association analysis in words and mathematically

Manhattan plot and explanation - explain what you mean by 'adapted threshold'

Below 5% (not 5)

lambda analysis (***including*** mathematical definition of lambda and **SQUARE** QQ plots) — incomplete - what is denom?

(*) principal (not 'principle')

5/8

put in a table

0.5/1

6. Identify significant SNPs (include all relevant information)

0.75/1

7. Plots:

label size (not too small) informative captions

placement explanations

0.75/1

8. Conclusions

(somewhat vague and unclear)

1. recap analysis 2. state and interpret main findings

0.5/1

9. Overall presentation (clarity of explanations, appropriate citations / references) :

many run-on sentences (throughout)

poor satisfactory good excellent

10. Other comments:

Breheny

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other:

2.5/4

Read

GWAS: Name NT 6/12 → 3/6

1. Formatting:

all margins 2.5cm	informative title
12 pt size	name on all pages
no raw R code or output	all pages numbered
max **10** pages	no blurry plots (**NOT png**)

0.75/1

2. Introduction/Background:

brief background and statement of scientific question – why CAD (since you don't analyze it)

all variables defined

0.75/1

3. EDA / PCA:

relevant histograms / exploratory plots (NO BOXPLOTS)

EXPLAIN RELATION between PCs and population stratification

plot pc2 (y-axis) vs pc1 (x-axis)

0.75/2 + do you do pc with corr or cov matrix? (square) Did you consider that the scale of deviation is small so no real clusters

4. Pre-processing / QC steps: CLEARLY EXPLAIN

0.75/2 use sectioning to clarify

SNP QC: criteria and reasons

sample QC: criteria and reasons unclear

Hardy-Weinberg equilibrium: DEFINE and say how it relates to quality all pts

measure of LD and how it is used – explain not defined + Devlin+Roeder

5. Association / post-association analysis: + give PRIMARY references

write out model ***mathematically*** (for a given SNP): CLEARLY define all variables and domains - Model MUST RELATE TO SNP + what is SRK?

0.75/2

describe association analysis in words and mathematically what hypo test?

Manhattan plot and explanation (incomplete - explain thresholds don't just state)

lambda analysis (***including*** mathematical definition of lambda and **SQUARE** QQ plots) + EXPLANATION incomplete – what is denom?

3.75/8

0.5 put results in a table

6. Identify significant SNPs (include all relevant information)

0.75 7. Plots:

label size (not too small) informative captions

placement explanations

0.5 8. Conclusions use paragraphing not 'ensured reliability'

1. recap analysis clearly 2. state and interpret main findings

0.5 9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: _____

_____

_____

_____

_____

_____

2-25/4