GWAS: Name JD 7/12 → 3.5/6

Redd

1. Formatting: 0.75/0.75

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**) – too many digits (λ)

2. Introduction/Background: 0.5/1

brief statement of scientific question

all variables defined

3. PCA: 0.5/2

explain relation between PCs and population stratification

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

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

SNP QC: criteria and reasons

sample QC: criteria and reasons – be explicit

Hardy-Weinberg equilibrium: what it means and how it relates to quality

Overall QC explanation – very superficial, explain reasons

5. Association / post-association analysis: 1.25/2

Describe association analysis in words and mathematically → more usual to write $\epsilon \sim N(0, \sigma^2)$ & no ^ if have error term and leave σ out of lin model

Manhattan plot → x-axis: chromosomal location, y-axis: $-\log_{10}$ P-value } explain in text (clearly)

lambda analysis (including **SQUARE** QQ-normal plots) → not defined

LD heatmap (optional – does NOT count); measure of LD

3.5/7.75

0.75/1

6. Write out final estimated model **mathematically** (for a given SNP)

^ everywhere in est model, or nowhere with error term

hat on response variable

MUST RELATE TO SNP

0.75/1

7. Plots:

label size (not too small)

captions

placement

NOT BLURRY

1/1

8. Conclusions

a little brief

recap analysis

somewhat generic but OK

state main findings

1/1

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

poor satisfactory good excellent

10. Other comments:

- Don't need R object/plink details
- Don't need table of contents format, just write in line

3.5/4.25

Comments **Name:** JD

Reed

00 – informative title

A – eda

B – PCA + explain

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically* ^ everywhere or nowhere

I – Manhattan plot (and explanation)

J – identify significant markers

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text)

R – overall organization and explanation of procedure superficial

S – other:

- |{min. alleles at jth Snp}| – just write: number of minor alleles at SNP j

(A) Don't need to re-do

great job!!

GWAS: Name Reed GD 11.25/12 → 5.625/6 → 6/6

1. Formatting:

0.75/0.75

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 + .25 / 1

2. Introduction/Background:

brief statement of scientific question

all variables defined

1.75/2

3. PCA:

explain relation between PCs and population stratification

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

2/2

4. Pre-processing / QC steps: excellent!

SNP QC: criteria and reasons

sample QC: criteria and reasons

Hardy-Weinberg equilibrium: what it means and how it relates to quality

Overall QC explanation

1.5/2

5. Association / post-association analysis:

Describe association analysis in words and mathematically ↑ Z-test, not anova

Sex not defined ^ everywhere or nowhere

Manhattan plot $-\log_{10} P$

Define lambda analysis (including **SQUARE** QQ-normal plots)

LD heatmap (optional – does NOT count); measure of LD

7.25/7.75

1/1

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP

1.25/1.25

7. Plots:

label size (not too small) captions

placement **NOT BLURRY**

0.5/1

8. Conclusions use paragraphs

incomplete

recap analysis

(*) interpretation

state main findings

1.25/1

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

poor satisfactory good (excellent)

10. Other comments:

(*) cannot conclude causation, only association

4/4.25

Comments **Name:** GD

Read

00 – informative title

A – eda

B – PCA + explain

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically* – define sex

I – Manhattan plot (and explanation)

J – identify significant markers

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text)

R – overall organization and explanation of procedure

S – other:

great job !! ☺

GWAS: Name _____ FD 6.25/12 → 3.125/6

Reed

1. Formatting:

0.5/0.75

- 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**)

2. Introduction/Background:

0.75/1

- brief statement of scientific question
- all variables defined

- too many digits
- imprecise - cannot conclude causality, only association

0.5/1

3. PCA:

- explain relation between PCs and population stratification
- plot pc2 (y-axis) vs pc1 (x-axis)

- SQUARE

4. Pre-processing / QC steps:

Be clear and specific

0.5/2

- SNP QC: criteria and reasons
- sample QC: criteria and reasons
- Hardy-Weinberg equilibrium: what it means and how it relates to quality
- Overall QC explanation

define

- somewhat superficial

5. Association / post-association analysis:

What is the test?

0.75/2

- Describe association analysis in words and mathematically
- Manhattan plot
- lambda analysis (including **SQUARE** QQ-normal plots)
- LD heatmap (optional – does NOT count); measure of LD

x-axis: chromosomal location
y-axis: $-\log_{10}$ p-value

not defined

3/7.75

1/1 very good

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP

0.75/1.25 7. Plots: some plots very small

label size (not too small) captions - number + caption each figure

(placement) - blank space **NOT BLURRY**

0.75/1 8. Conclusions

recap analysis expand on this state main findings

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

poor satisfactory good excellent

10. Other comments:

- refs incomplete + cite in text
- explain using your own words

3.25/4.25

Comments **Name:** FD

Read

00 – informative title

A – eda

B – PCA + explain

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically*

I – Manhattan plot (and explanation)

J – identify significant markers

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions brief

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text)

R – overall organization and explanation of procedure

S – other:

_____

_____

_____

_____

_____

GWAS: Name _____ MD 9.5/12 → 4.75/6

Read

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**)

0.75/0.75

2. Introduction/Background:

brief statement of scientific question (imprecise but ok)

all variables defined – Don't need files

1/1

3. PCA: +explain

explain relation between PCs and population stratification

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

0.5/2

4. Pre-processing / QC steps: – Did you create this schema?

SNP QC: criteria and reasons incomplete detail

sample QC: criteria and reasons detail – be specific

Hardy-Weinberg equilibrium: what it means and how it relates to quality

Overall QC explanation

Don't 'ensure' adherence

1.5/2

5. Association / post-association analysis:

Describe association analysis in words and mathematically → if est. coef (not 'model') p-value < thresh
↳ sex not defined

Manhattan plot

lambda analysis (including **SQUARE** QQ-normal plots) Define

LD heatmap (optional – does NOT count); measure of LD

(OK)

5.25/7.75

0.75/1

sex not defined

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP

1.25/1.25

7. Plots:

label size (not too small) captions

placement **NOT BLURRY**

1/1

8. Conclusions

don't 'validate' tutorial results

recap analysis

state main findings

1.25/1

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

poor satisfactory good excellent

10. Other comments:

- cite primary refs (not course notes [4])

- good job !! :)

4.25/4.25

Comments **Name:** MD

Red

00 – informative title

A – eda

B – PCA + explain

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically*

I – Manhattan plot (and explanation)

J – identify significant markers

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text)

R – overall organization and explanation of procedure

S – other:

- Define all terms (minor allele, etc)

GWAS: Name _____ 7.5/12 → 3.75/6

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**)

2. Introduction/Background:

brief statement of scientific question

all variables defined

3. PCA:

explain relation between PCs and population stratification

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

4. Pre-processing / QC steps:

SNP QC: criteria and reasons

sample QC: criteria and reasons

Hardy-Weinberg equilibrium: what it means and how it relates to quality

Overall QC explanation

5. Association / post-association analysis:

Describe association analysis in words and mathematically

Manhattan plot

lambda analysis (including **SQUARE** QQ-normal plots)

LD heatmap (optional – does NOT count); measure of LD

Table of top sig results

4/7.75

1/1 (no ^ if $+\epsilon$) (ok given previous)

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable — MUST RELATE TO SNP

0.75/1.25

7. Plots:

label size (not too small) — captions number + descriptive caption for each figure

placement — **NOT BLURRY**

0.75/1

8. Conclusions

don't 'verify' model

recap analysis — state main findings

Ghard to follow in parts

1/1

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

poor — satisfactory — good — excellent

But: be careful to use your own words, some of your explanations appear to be unattributed quotes

10. Other comments:

- cite refs in-text, no 'general' refs
- refs incomplete

3.5/4.25

Comments **Name:** A Duf

Reed

00 – informative title

A – eda

B – PCA + explain

C – define and explain HWE mathematically

D – define and explain HWE **test**

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically* no ^ if te

I – Manhattan plot (and explanation) fix typo

J – identify significant markers Results table

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions (could be more clear)

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text)

R – overall organization and explanation of procedure

S – other:

- Don't use footnotes, it's distracting
- cite primary refs (not wikipedia, moodle, etc)

GWAS: Name _____ LF 6/12 → 3/6

Reed

1. Formatting:

0.75/0.75

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:

1/1

brief statement of scientific question

all variables defined

0.25/2

3. PCA: – explain

+ explain relation between PCs and population stratification

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

4. Pre-processing / QC steps:

0.5/2

SNP QC: criteria and reasons

sample QC: criteria and reasons

Hardy-Weinberg equilibrium: what it means and how it relates to quality — incomplete definition

Overall QC explanation — superficial

5. Association / post-association analysis:

0.75/2

Describe association analysis in words and mathematically ^ everywhere or nowhere

Manhattan plot + explain

lambda analysis (including **SQUARE** QQ-normal plots) — Define

LD heatmap (optional – does NOT count); measure of LD — snp, sex not defined

– Table of top sig results

3.25/7.75

0.5/1 ↑ everywhere or nowhere

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP (define Snp, Sex

0.75/1 (.25)

7. Plots:

label size (not too small) captions

placement **NOT BLURRY**

0.5/1

8. Conclusions — no goal not to 'do gwas' — gwas is a tool for identifying associations in the data

recap analysis (brief) state main findings — expand

1/1

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

poor satisfactory good excellent

10. Other comments:

- Dont need cover page / course name / EPFL logo

- you have several unattributed quotes, use your own words

2.75/4.25

Comments **Name:** LF

Read

00 – informative title

A – eda

B – PCA + explain

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically* sex, snp not defined

I – Manhattan plot (and explanation)

J – identify significant markers Table of results

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M – interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text) Blank Space

R – overall organization and explanation of procedure

S – other:

- Define all terms (minor allele, etc)

GWAS: Name _____ AH 4.5/12 → 2.25/6

Read

1. Formatting: 0.5/0.75

~~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: 0.5/1 imprecise
   - brief statement of scientific question — be specific
   - all variables defined

3. PCA: 0.25/2 explanation imprecise and not completely correct
   - explain relation between PCs and population stratification
   - plot pc2 (y-axis) vs pc1 (x-axis) + SQUARE

4. Pre-processing / QC steps: 0.5/2
   - SNP QC: criteria and reasons → you left out ? sign — don't only remove individuals, also SNPs
   - sample QC: criteria and reasons
   - Hardy-Weinberg equilibrium: what it means and how it relates to quality — all freqs → explain
   - ~~Overall QC explanation~~ very superficial and unclear
   - → Don't EXCLUDE high call rate !!

5. Association / post-association analysis: 0.5/2 be specific
   - Describe association analysis in words and mathematically
   - Manhattan plot — spurious explanation confusing and incorrect
   - lambda analysis (including **SQUARE** QQ-normal plots) + explain
   - LD heatmap (optional – does NOT count); measure of LD

→ x-axis: chromosomal location
→ y-axis: $-\log_{10}$ p-value
- suggestive threshold incorrect
- explain meaning of sig results in CETP

2.25/7.75

0.25/1

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP

↳ not in model + Sex not defined

0.75/1.25

7. Plots:

label size (not too small) captions

placement **NOT BLURRY**

0.25/1

8. Conclusions use paragraphs

recap analysis state main findings

vague in parts

1/1

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

poor satisfactory good excellent

10. Other comments:

- cite refs in text — no general refs

- much of what you say is generic (rather than specific) and imprecise

Example: The goal of assoc analysis is in general to carry out reg on each SNP.

NO: we don't have reg as a goal, it's a tool that can be used to identify associations in the data.

→ See the difference? (+ more like this throughout report)

2.25/4.25

Comments **Name:** AH

Read

00 – informative title

A – eda

B – PCA + explain

C – define and explain HWE – all gt freqs

D – define and explain HWE **test** more specifics

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically* incomplete

I – Manhattan plot (and explanation) imprecise

J – identify significant markers

K – square QQ plots

L – fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions vague in parts

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text)

R + overall organization and explanation of procedure

S + other:

- Define all terms (call rate / minor allele / etc)
- controls not defined
- QC doesn't 'ensure' anything

GWAS: Name AM 5.5/12 → 2.75/6

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:

brief statement of scientific question

all variables defined

3. PCA:

explain relation between PCs and population stratification

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

4. Pre-processing / QC steps:

SNP QC: criteria and reasons

sample QC: criteria and reasons

Hardy-Weinberg equilibrium: what it means and how it relates to quality

Overall QC explanation

5. Association / post-association analysis:

Describe association analysis in words and mathematically

Manhattan plot

lambda analysis (including **SQUARE** QQ-normal plots)

LD heatmap (optional – does NOT count); measure of LD

3/7.75

0.75/1

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP

0.5/1.25

7. Plots: + most plots too small

label size (not too small) captions – make informative

placement (**NOT BLURRY**)

0.25/1

8. Conclusions Too brief

recap analysis state main findings

1/1

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

poor satisfactory good excellent

10. Other comments:

- Don't need to cite R fns

2.5/4.25

Comments **Name:** AM

Read

00 – informative title

A – eda

B – PCA + explain

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically* (no ^ if +ϵ)

I – Manhattan plot (and explanation) y-axis label

J – identify significant markers Results table

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text)

R – overall organization and explanation of procedure

S – other:

- Define all terms (minor allele, etc
- use your own words, there appear to be several unattributed quotes
- Delete Figure 4, and anyway you don't cite it

GWAS: Name _____ SN 7.25/12 → 3.625/6

Read

1. Formatting:

0.75/0.75

- 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**)

0.5/1 2. Introduction/Background:

- brief statement of scientific question
- all variables defined

- too many digits
- too short + imprecise

0.5/2 3. PCA:

- explain relation between PCs and population stratification
- plot pc2 (y-axis) vs pc1 (x-axis)

- explanation: not your own words + be precise

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

- SNP QC: criteria and reasons
- sample QC: criteria and reasons
- Hardy-Weinberg equilibrium: what it means and how it relates to quality
- Overall QC explanation

be specific - your reasoning is vague and generic

What are the controls?

1/2 5. Association / post-association analysis:

- Describe association analysis in words and mathematically
- Manhattan plot
- lambda analysis (including **SQUARE** QQ-normal plots)
- LD heatmap (optional – does NOT count); measure of LD

→ regress Y on X

incomplete

- y-axis label → -log10 p-value + explain

HDL - 'related' (not 'linked') to CAD, because linkage has a very technical meaning in genetics

3.25/7.75 - Hists of cholesterol levels

+ $\hat{\beta}$ snp in results table

1/1 6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP

1.25/1.25 7. Plots:

label size (not too small) captions

placement **NOT BLURRY**

0.75/1 8. Conclusions

recap analysis state main findings

'validity' of models?

- last paragraph should be earlier

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

poor satisfactory (good) excellent

10. Other comments:

- see other page

4/4.25

Comments **Name:** SN

Reed

00 – informative title

A – eda

B – PCA + explain

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically*

I – Manhattan plot (and explanation)

J – identify significant markers

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text)

R – overall organization and explanation of procedure

S – other:

- cite Reed et al

GWAS: Name _____ WO 4.25/12 → 2.125/6

Reed

1. Formatting:

0/0.75

- 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**)

0.5/1 2. Introduction/Background:

- brief statement of scientific question — be specific
- all variables defined — explain in your own words

0.5/2 3. PCA: (+EDA) + explain technique

- explain relation between PCs and population stratification
- plot pc2 (y-axis) vs pc1 (x-axis) + SQUARE

4. Pre-processing / QC steps:

0.5/2

- SNP QC: criteria and reasons
- sample QC: criteria and reasons
- Hardy-Weinberg equilibrium: what it means and how it relates to quality
- Overall QC explanation — very generic and superficial → why? reason's — be specific

0.75/2 5. Association / post-association analysis:

- (*) Describe association analysis in words and mathematically — what hyp test ?? — SNP/sex not defined
- Manhattan plot y-axis: $-\log_{10}$ p-value
- lambda analysis (including **SQUARE** QQ-normal plots)
- LD heatmap (optional – does NOT count); measure of LD

- what is 'background of each case and control'?

(*) your interpretation of p-value is incorrect

- model has some typos
- Explain Post-assoc, what you have is very generic and non-specific

2.25

0.5/1 ^everywhere or nowhere

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP define snp/sex

0.75/1.25 7. Plots:

label size (not too small) captions make more informative

placement **NOT BLURRY**

0.5/1 8. Conclusions

recap analysis (**) interpretation

state main findings

0.25/1 9. Overall presentation (clarity of explanations, appropriate citations / references) : use a spell checker + use your own words (*)

poor satisfactory good excellent

10. Other comments:

(*) several apparently unattributed quotes

(**) cannot conclude causation, only association

2/4.25

Comments **Name:** WD

Need

00 – informative title

A – eda

B – PCA + explain

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure (use own words)

G – explain association test mathematically)

H – write out final model *mathematically* not entirely correct)

I – Manhattan plot (and explanation)

J – identify significant markers

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text) too small

Q – plot layout (see text)

R – overall organization and explanation of procedure *

S – other:

define all terms: minor allele, etc

* use your own words

- inbreeding coef $ < $

- no refs

- what you have above conclusion is very confusing

(*) Don't need to re-do good job!! :)

GWAS: Name Reed ST 10/12→5/6→6/6

1. Formatting:

0.75/0.75

- 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**)

1/1 +.25

2. Introduction/Background:

- brief statement of scientific question
- all variables defined

0.75/2

3. PCA: somewhat incomplete

- explain relation between PCs and population stratification
- plot pc2 (y-axis) vs pc1 (x-axis) - SQUARE

4. Pre-processing / QC steps:

1.5/2

- SNP QC: criteria and reasons — more specifically
- sample QC: criteria and reasons
- Hardy-Weinberg equilibrium: what it means and how it relates to quality — implied - be explicit
- Overall QC explanation — very clear!

5. Association / post-association analysis:

1.25/2

excellent

- Describe association analysis in words and mathematically
- Manhattan plot — x-axis: chromosomal location; y-axis: $-\log_{10}$ p-value
- lambda analysis (including **SQUARE** QQ-normal plots)
- LD heatmap (optional – does NOT count); measure of LD

- regress Y on X
- reduce (not "minimize") impact of confounders — control for (potential) confounders 5
- give small table of most sig results

5.5/7.75

0.75/1 ^ everywhere (all est coefs)

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP

0.75/2 7. Plots:

label size (not too small) captions

placement **NOT BLURRY**

1 + .25 / 1 8. Conclusions

recap analysis state main findings

1.25/1 9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

+.5 pheeras

4.5/4.25

Comments **Name:** ST

Reed

00 – informative title

A – eda

B – PCA + explain

C – define and explain HWE

D – define and explain HWE **test** more explicitly

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically*

I – Manhattan plot (and explanation)

J – identify significant markers Results table

K – square QQ plots

L – fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text)

R – overall organization and explanation of procedure

S – other:

great job !!

GWAS: Name Reed AT 6/12 → 3/6

1. Formatting: 0.75/0.75

- 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**)

2. Introduction/Background: 0.75/1

- brief statement of scientific question — ok but a little unclear
- all variables defined

→ Break into more paragraphs

3. PCA: 0.5/2 — no EDA

- explain technique
- explain relation between PCs and population stratification
- plot pc2 (y-axis) vs pc1 (x-axis) + SQUARE

4. Pre-processing / QC steps: don't 'ensure' quality

0.5/2

- SNP QC: criteria and reasons
- sample QC: criteria and reasons
- Hardy-Weinberg equilibrium: what it means and how it relates to quality
- Overall QC explanation

write mathematically

very superficial, you need to explain reasons, not state conclusions

Explain

5. Association / post-association analysis: 0.5/2

- Describe association analysis in words and mathematically
- Manhattan plot + explain
- lambda analysis (including **SQUARE** QQ-normal plots)
- LD heatmap (optional – does NOT count); measure of LD

y-axis: $-\log_{10}$ p-value

test?

- sex/age not for substructure

define

- Table of most sig results

all very spread out, hard to follow; I need to search around for the info

3/7.75

0.75/1 ^ on all est coefs

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP

0.75/1

7. Plots:

label size (not too small) captions

placement **NOT BLURRY**

0.75/1

8. Conclusions use paragraphs atherosclerosis or CAD?

recap analysis state main findings very vague/general

0.75/1

9. Overall presentation (clarity of explanations, appropriate citations / references) : very hard to follow

poor satisfactory good excellent

10. Other comments:

- make shorter paragraphs - hard to follow

3/4.25

Comments **Name:** AT

Read

00 – informative title

A – eda

B – PCA + explain

C – define and explain HWE mathematically

D – define and explain HWE **test**

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically* ↑ on all est. coefs

I – Manhattan plot (and explanation)

J – identify significant markers) Table of results

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text) put in text, not at end

R – overall organization and explanation of procedure

S – other:

- use sectioning
- include plots where they are discussed, not at end
- needs more paragraphing for clarity
- Define all terms – call rate, MAF, etc

GWAS: Name _____ ZW 5/12 → 2.5/6

Reed

1. Formatting:

0.5/0.75

- 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**)

0.5/1

2. Introduction/Background: use paragraphs

- brief statement of scientific question — be specific
- all variables defined + eda

0.25/2

3. PCA:

- explain relation between PCs and population stratification
- plot pc2 (y-axis) vs pc1 (x-axis) + SQUARE

0.25/2

4. Pre-processing / QC steps:

- SNP QC: criteria and reasons
- sample QC: criteria and reasons
- Hardy-Weinberg equilibrium: what it means and how it relates to quality
- Overall QC explanation — very superficial and incomplete

5. Association / post-association analysis:

0.75/2

- Describe association analysis in words and mathematically — everywhere or nowhere
- Manhattan plot x-axis: chromosomal location — test? — sex not defined
- lambda analysis (including **SQUARE** QQ-normal plots)
- LD heatmap (optional – does NOT count); measure of LD
- why inverse normal transform?
- show graphical evidence
- what is your R code for this plot?
- Table of results

2.25/7.75

0.5/1 all ^ or none define sex

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP

0.75/1.25 7. Plots: some plots too small

label size (not too small) captions

placement **NOT BLURRY**

0.5/1 8. Conclusions use paragraphs

recap analysis (rather brief) state main findings expand

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

poor satisfactory good excellent

10. Other comments:

2.75/4.25

Comments **Name:** ZW

Read

00 – informative title

A – eda

B – PCA + explain

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically* sex not defined

I – Manhattan plot (and explanation)

J – identify significant markers Results table

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text)

R – overall organization and explanation of procedure

S – other:

– use your own words