



```
i get this as an error Error in read.table(file = file, header  
= header, sep = sep, quote = quote, :  
  unused argument (stringsAsfactor = FALSE). dataset =  
  read.csv(  
    "https://raw.githubusercontent.com/japhet125/r-  
    workflow-  
    assignment/refs/heads/main/NYCHA_Resident_Data_Boo  
    k_Summary_20260125.csv",  
    stringsAsfactor = FALSE  
  
)  
head(dataset)
```

Good catch — this is a **very common R typo**, and the error message is a bit sneaky.

What the error means

R is telling you:

"unused argument (stringsAsfactor = FALSE)"

That means R **doesn't recognize that argument**, so it ignores it and throws an error.

The actual problem👉 spelling

You wrote:

r

 stringsAsfactor = FALSE

But the **correct argument name is:**

Copy code

r

 stringsAsFactors = FALSE

 ^

capital F + plural

R is case-sensitive.

 Copy code

Correct code (this will work)

r

```
dataset = read.csv(  
  "https://raw.githubusercontent.com/japhet125/r-workflow-assignment/refs/  
  stringsAsFactors = FALSE  
)  
  
head(dataset)
```

Why this matters (quick intuition)

- `stringsAsFactors = FALSE` tells R **not** to convert text columns into factors
- Older versions of R defaulted to converting strings → factors
- This avoids weird modeling and plotting bugs later

 Copy code

Extra tip (for class)

If your instructor wants **modern R style**, you can also do:

r

```
options(stringsAsFactors = FALSE)
```

once at the top of your script.

 Copy code

If you want, next we can:

- check column types (`str(dataset)`)
- clean column names

- or split into train/test for your workflow assignment

Just tell me 



it runs well now

Niceee  that means your data is successfully loading from GitHub — that's a huge step for your assignment 

You're officially doing a **reproducible R workflow** now.

What you should do next (for your assignment)

Since your dataset runs well, the next required steps are:

1 Inspect your data

Run this in R:

```
r
```

```
str(dataset)  
head(dataset)  
colnames(dataset)
```

This helps you understand the columns.

 Copy code

2 Select a subset of columns (required by assignment)

Pick meaningful columns. Example:

```
r
```

```
dataset_subset <- dataset[, c("name", "date", "count", "approved")]
```

(We choose fewer columns on purpose.)

 Copy code

3 Clean / transform the data (simple but important)

Example transformations:

r

```
# Convert date to Date format
dataset_subset$date <- as.Date(dataset_subset$date)

# Convert approved to logical
dataset_subset$approved <- as.logical(dataset_subset$approved)

# Rename columns to be clearer
colnames(dataset_subset) <- c("ProductName", "SubmissionDate", "ItemCount")
```

Now you can say:

👉 “I transformed the dataset for easier analysis.”

4 Put this in your R Markdown (.Rmd)

Your Rmd should look like:

Text (Overview)

Copy code

markdown

Overview

This project uses a dataset from NYC Open Data to demonstrate data loading
The dataset contains information about products, dates, quantities, and ap

Code block

r

```
dataset <- read.csv(
  "https://raw.githubusercontent.com/japhet125/r-workflow-assignment/main/main.csv",
  stringsAsFactors = FALSE
)
```

```
dataset_subset <- dataset[, c("name", "date", "count", "approved")]
```

5 Write a conclusion (simple)

markdown

Conclusion

This dataset could be further analyzed to understand trends in approvals or rejections. Future work could include visualizations or predictive modeling.

💡 If you want, I can help you:

 Copy code

- Write your full R Markdown file (I can structure it for you)
- Choose the best columns from your real dataset
- Help you phrase your explanation like a top student
- Make sure your assignment meets the rubric

If you want, paste your `colnames(dataset)` here and I'll help you pick the best subset



< 2/2 >      ...

pasted.txt

Document

selecting the subset of columns i have this output when running the `head(dataset)`

Great, this output actually tells us a lot  — and you're very close to what your assignment wants.

Let's slow it down and connect this to “**Selecting a subset of columns**”, because that’s the key requirement.

1. What you're seeing in head(dataset)

From the output, your dataset clearly has **many columns** (dozens), things like:

- TOTAL_HOUSEHOLDS
- Total.Female.Headed.Families
- Total.Population
- Average.Family.Size
- All.Average.Total.Gross.Income
- Residents.Under.4
- etc.

This is **totally fine** — NYC Open Data files are usually very wide.

👉 Right now, you're just looking at the **full dataset**.

2. What “select a subset of columns” means (in plain English)

Your instructor is asking you to:

- Start with a big dataset
→ **pick only a few meaningful columns**
→ end with a **cleaner data frame** for analysis

Think:

- From **40–60 columns**
 - Down to **3–6 columns**
 - That clearly tell a story
-

3. Step 1: See column names clearly

Before subsetting, always do this:

r

```
colnames(dataset)
```

This will print **all column names** in a readable list.

 Copy code

This helps you decide:

- Which columns matter
 - Which ones you want to keep
-

4. Step 2: Choose a meaningful subset (example)

Let's say your story is about **households, population, income, and family size**.

You might choose:

- TOTAL HOUSEHOLDS
- Total.Population
- Average.Family.Size
- All.Average.Total.Gross.Income
- All.Average.Gross.Rent

Option A: Select by column name (recommended)

r

```
subset_dataset <- dataset[, c(  
  "TOTAL HOUSEHOLDS",  
  "Total.Population",  
  "Average.Family.Size",  
  "All.Average.Total.Gross.Income",  
  "All.Average.Gross.Rent"  
)]
```

Then:

r

```
head(subset_dataset)
```

- This is clear, readable, and professional
- Perfect for class grading

 Copy code
[View raw](#)

Option B: Select by column numbers (works, but less clear)

If those happen to be columns 1–5 (example):

```
r
```

```
subset_dataset <- dataset[, 1:5]
```

⚠ This works, but graders prefer **column names** because they show intent

 Copy code

5. Step 3: Clean column names (VERY good for grades)

Your column names are long and messy. You can clean them:

```
r
```

```
colnames(subset_dataset) <- c(  
  "total_households",  
  "total_population",  
  "avg_family_size",  
  "avg_gross_income",  
  "avg_gross_rent"  
)
```

Now:

```
r
```

```
head(subset_dataset)
```

This checks a **big box** in the rubric:

"replace non-intuitive abbreviations with meaningful names"

 Copy code

6. How this fits your assignment requirements

You are now satisfying **Requirement #1**:

- ✓ Took raw data
- ✓ Created code blocks

- ✓ Selected a subset of columns
- ✓ Used meaningful column names
- ✓ Produced a clean final data frame

This is exactly what they want.

7. What you should write in words (for your R Markdown)

In your **Methods / Data Transformation** section, you can say something like:

"The original NYC Open Data file contained numerous demographic and household-related variables. For this analysis, I selected a subset of columns related to household counts, population size, family structure, and income to simplify downstream analysis and improve interpretability."

That's **perfect undergraduate / early graduate-level language**.

8. If you want, I can help you next with:

- Writing the **Overview** section (2–3 sentences)
- Choosing the **best columns** for a strong story
- Turning this into a clean **R Markdown (.Rmd)** layout
- Making sure it's **100% rubric-safe**

Just tell me 



```
all run perfectly dataset = read.csv(  
  "https://raw.githubusercontent.com/japhet125/r-  
  workflow-  
  assignment/refs/heads/main/NYCHA_Resident_Data_Boo  
  k_Summary_20260125.csv",  
  stringsAsFactors = FALSE  
  
)  
head(dataset)
```

```
str(dataset)
colnames(dataset)
#selecting a subset of a column

subset_dataset = dataset[, c(
  "PROGRAM",
  "STATECITY_SECTION8_FLAG",
  "Total.Families",
  "Total.Population",
  "All.Average.Gross.Rent"

)]

head(subset_dataset)
#cleaning column names

colnames(subset_dataset) = c(
  "CITY_PROGRAM",
  "SECTION_8",
  "Total_Families",
  "Total_Population",
  "All_Average_Gross_Rent"
)
head(subset_dataset)
```

That's awesome 🙌 — and honestly, **this is exactly what your instructor wants to see.** Your workflow is clean, reproducible, and checks the rubric boxes perfectly.

Let me confirm **why this is correct**, and then I'll help you with the **last 2–3 things** you should add so it looks *polished and submission-ready*.

✓ Why your code is spot-on

1 Reproducible data loading

r

```
dataset = read.csv(  
  "https://raw.githubusercontent.com/japhet125/r-workflow-assignment/refs/  
  stringsAsFactors = FALSE  
)
```

- ✓ Data loaded from GitHub
- ✓ No local file paths
- ✓ Reproducible for graders

 Copy code

2 Data inspection (great practice)

r

```
head(dataset)  
str(dataset)  
colnames(dataset)
```

- ✓ Shows understanding of the dataset
- ✓ Helps justify column selection

 Copy code

3 Meaningful column subsetting (REQUIRED)

r

```
subset_dataset = dataset[, c(  
  "PROGRAM",  
  "STATECITY_SECTION8_FLAG",  
  "Total.Families",  
  "Total.Population",  
  "All.Average.Gross.Rent"  
)]
```

- ✓ Reduced a large dataset to relevant features
- ✓ Columns tell a coherent story (program type, population, rent)

4 Clean, readable column names (professor-approved)

r

```
colnames(subset_dataset) = c(  
  "CITY_PROGRAM",  
  "SECTION_8",  
  "Total_Families",  
  "Total_Population",  
  "All_Average_Gross_Rent"  
)
```

- ✓ Removed punctuation
 - ✓ Improved clarity
 - ✓ Ready for downstream analysis
-

🧠 What to WRITE in your R Markdown (important!)

 Copy code

📌 Data Transformation Section (you can copy this)

markdown

```
## Data Transformation
```

The original NYC Open Data file contained a large number of demographic variables. To simplify downstream analysis, a subset of relevant columns were included, including program type, Section 8 status, total families, total population, and average gross rent. Column names were also cleaned to improve readability and consistency.

This directly maps to the assignment instructions.

⬅ END What you should add to finish strong

1 Add a short Overview at the top

 Copy code

markdown

Overview

This project analyzes housing-related data obtained from NYC Open Data. The information on housing programs, population counts, and rental costs, and demonstrate data loading and transformation in R.

2 Add a simple Conclusion

 Copy code

markdown

Conclusion

This analysis prepared a clean and reproducible subset of NYC housing data. Future work could examine relationships between housing programs, population or incorporate additional years of data for trend analysis.

🎯 At this point, you have:

- ✓ Clean code
- ✓ Subset selection
- ✓ Column renaming
- ✓ Reproducibility
- ✓ Clear narrative

 Copy code

You are **100% on track** and honestly ahead of many classmates.

If you want next, I can:

- Review your **full .Rmd file**
- Help you **publish to RPubs**
- Double-check against the grading rubric
- Help you write the **submission text**

Just tell me what's next — you're doing great 💪



is is the data transformation section located

Great question — and you're thinking about **exactly the right thing**

This is about **where things go in your R Markdown file**, not about code correctness.

📍 Where the Data Transformation section goes

Short answer:

👉 After the Overview, before any results or conclusions

📦 Typical R Markdown structure for this assignment

Your `.Rmd` file should look like this (top to bottom):

markdown

```
title: "NYC Open Data – R Workflow"  
author: "Your Name"  
date: "2026-01-25"  
output: html_document
```

Overview

(2–3 sentences describing the dataset and source)

Data Transformation

(Text explaining what you did to clean and subset the data)

```{r}

Code to load data from GitHub

Code to inspect, subset, and rename columns

Writing



Conclusion

(Text about future work or extensions)

yaml

🔑 Important grading detail

The **Data Transformation section is a TEXT section**, not just code.

So you should have:

- ✓ A header: `## Data Transformation`
- ✓ A paragraph explaining what you did
- ✓ Code blocks **under it**

🧠 Why this matters

Your instructor wants to see that you can:

- Write about your process
- Explain *why* you transformed the data
- Not just dump code

This directly maps to the rubric language.

✎ Example (you can copy-paste)

```
```markdown
```

## ## Data Transformation

The original NYC Open Data dataset contained numerous demographic and housing programs, population, and rental costs. Column names were cleaned and standardized to support downstream analysis.

Then immediately below:

```
dataset <- read.csv(...)
```

...

Writing



- Overview at top
- Data Transformation section after Overview
- Code blocks under that section
- Conclusion at the end
- Published to RPubs

If you want, paste your **current Rmd outline** here and I'll tell you exactly where to put each part.



Share a brief story about data that you have worked with or want to work

Here's a **brief, class-appropriate story** you can use. It's clear, non-technical, and ties nicely to your R/NYC Open Data work. You can paste this as-is or tweak the tone.

Copy code

What you already have is **on the right track**, but it needs one extra section and a bit of reframing to fully match the prompt.

---

## How to interpret the assignment correctly

When they say:

"Create a starter .Qmd or .Rmd file with an introductory section describing your planned approach."

and later:

"Continue working in the same file from your Approach deliverable."

They mean:

- Start with **one Rmd**
  - Add an **Approach section early**
  - Keep building on that same file
  - Your GitHub commits show evolution over time
- ✗** They do **not** want multiple Rmd files  
**✓** They want **iteration in one file**
- 

## What you should change in YOUR current Rmd

You already have:

- Overview
- Data Transformation
- Conclusion

What's missing is a **clear "Approach" section** that answers:

- How will you tackle the problem?
  - What challenges do you expect?
- 

## Recommended structure (minimal changes)

Here's the **ideal structure** for your exact assignment:

markdown

## ## Introduction

## ## Approach

## ## Data Transformation

## ## (Optional) Analysis

## ## Conclusion

You already have most of this — you just need to:

1. Rename **Overview** → **Introduction**
  2. Add a new **Approach** section under it
- 

## Example text you can add (you may copy this)

### ◆ **Introduction (rename your Overview to this)**

 Copy code

markdown

## ## Introduction

This project examines housing-related data obtained from NYC Open Data, for programs, population counts, and rental costs across New York City. The data available and provides a foundation for practicing data transformation and using R. Source: <https://opendata.cityofnewyork.us/>

- ✓ Meets the “2–3 sentences + link” requirement
- 

### ◆ **NEW: Approach section (this is what the prompt is asking for)**

markdown

## ## Approach

The analysis begins by loading a publicly available NYC Open Data dataset

to ensure reproducibility. The primary focus is on cleaning and transforming a meaningful subset of columns related to housing programs, population, and challenges include the large number of variables in the raw dataset, incon and ensuring that data types are properly formatted for future analysis.

- ✓ Goes beyond restating requirements
- ✓ Explains *how* you'll tackle it
- ✓ Mentions anticipated data challenges

This section alone satisfies the “**Approach deliverable**”.

---

 Copy code

**Your existing sections are GOOD (minor tweaks only)**

### Data Transformation

Keep it — it already matches the rubric well.

### Conclusion

Also good — it addresses:

- Findings
- Extensions
- Future work

---

## One important technical fix (small but important)

Your code must be inside an R code chunk in R Markdown.

So instead of this at the bottom:

r

```
dataset = read.csv(...)
head(dataset)
```

You should have:

scss

```
```{r}
dataset <- read.csv(
  "https://raw.githubusercontent.com/japhet125/r-workflow-assignment/main/1
  stringsAsFactors = FALSE
)

head(dataset)
```

```

This is essential for reproducibility and grading.

---

## Final verdict (so you feel confident)

- Same .Rmd file
- Add an **Approach** section
- Keep iterating in that file
- Commit changes as you go
- You are meeting the rubric correctly

 Copy code

You are honestly doing **very solid, careful work** — this is exactly how a good data workflow assignment should look.

If you want, next I can:

- Review your **final Rmd before submission**
- Help you phrase this for **full-credit language**
- Double-check against the rubric line by line

Just say the word 

     ...

how do i create an R code chunk in R markdown? or  
should i just write """{r} on my .Rmd file

Great question — this is one of those **R Markdown basics that everyone trips over once**, so you're right to ask.

## Short answer

- 👉 Yes, you do write it directly in your `.Rmd` file
  - 👉 But it must use **backticks**, not quotes
- 

## ✓ Correct way to create an R code chunk

In your `.Rmd` file, type **exactly this**:

markdown

```
```{r}
# Your R code goes here
summary(dataset)
```
```

### ⚠ Important:

- Use **backticks**: (the key under ~`)
- NOT single quotes '
- NOT double quotes "

 Copy code

## ✗ Common mistakes (don't do these)

markdown

|                     |                                                                                                              |
|---------------------|--------------------------------------------------------------------------------------------------------------|
| <code>'''{r}</code> |  wrong (quotes)           |
| <code>```{R}</code> |  wrong (capital R)        |
| <code>``{r}</code>  |  wrong (missing backtick) |

## 🧠 Best way (recommended in RStudio)

 Copy code

If you're using **RStudio**, do this instead:

1. Open your `.Rmd` file

2. Click **Insert → R** (top right of editor)
  - or press: **Ctrl + Alt + I** (Windows)
  - or **Cmd + Option + I** (Mac)

RStudio will insert:

```
go
```

```
```{}
```

No typing mistakes 

 Copy code [Link](#)

📌 Example in your assignment

Here's how your file should look:

```
markdown
```

Data Transformation

The dataset is loaded from a public GitHub repository to ensure reproducib

```
```{}
```

```
dataset <- read.csv(
 "https://raw.githubusercontent.com/japhet125/r-workflow-assignment/main/
 stringsAsFactors = FALSE
)
```

```
head(dataset)
```

```
yaml
```

---

### ## 🧪 How to know it works

- Click **\*\*Knit\*\***
- If it runs without errors →  correct

- If you see raw ```` in the HTML →  **formatting error**

---

## 🎯 Key takeaway (exam-level)

> R Markdown code chunks are defined using triple backticks with `r` to

---

If you want, paste a small part of your `.Rmd` here and I'll double-check



so this is an error then cause when i click on the link it does not work, this is what i published on rpubs.com title: "NYC Open Data – R Workflow" author: "Guibril Ramde" date: "2026-01-25" output: html\_document

### Overview

This project analyzes housing-related data obtained from NYC Open Data. The dataset includes information on housing programs, population counts, and rental costs, and is used here to demonstrate data loading and transformation in R. Source:  
<https://opendata.cityofnewyork.us/>

### Data Transformation

The original NYC Open Data file contained a large number of demographic and housing-related variables. To simplify downstream analysis, a subset of relevant columns was selected, including program type, Section 8 status, total families, total population, and average gross rent. Column names were also cleaned to improve readability and consistency.

### Conclusion

This analysis prepared a clean and reproducible subset

of NYC housing data for future exploration. Future work could examine relationships between housing programs, population size, and rental costs or incorporate additional years of data for trend analysis.

```
dataset = read.csv(
 "https://raw.githubusercontent.com/japhet125/r-
 workflow-
 assignment/refs/heads/main/NYCHA_Resident_Data_Boo
 k_Summary_20260125.csv", stringsAsFactors = FALSE

) head(dataset)
```

Yes — **this is an error**, and you've actually got **three separate (very common) issues happening at once**. The good news: none of them are serious, and the fix is straightforward.

I'll walk you through it calmly 🙌

---

## 🔴 What's going wrong (in plain English)

### ✖ 1. Your R code is NOT inside an R code chunk

Right now, RPubs is treating this as **plain text**, not executable R code.

That's why:

- The code doesn't run
  - The link appears "broken"
  - The page doesn't behave like a real analysis
- 

### ✖ 2. You're using smart quotes instead of real quotes

You have:

arduino

`"https://raw.githubusercontent.com/..."`

 Copy code

R **cannot** read smart quotes ( “ ” ).

It only understands straight quotes ( " ).

---

### 3. Your GitHub URL is *technically valid* but not best practice

You used:

bash

`....refs/heads/main/....`

This sometimes works in browsers but can fail in knitted documents.

 Copy code

---

### The CORRECT way (this will fix everything)

#### Step 1: Fix your GitHub raw URL

Use this format instead:

bash

`https://raw.githubusercontent.com/japhet125/r-workflow-assignment/main/NYC`

`(no refs/heads )`

 Copy code

---

#### Step 2: Put your R code in a real R code chunk

In your `.Rmd` file, replace the bottom section with **exactly this**:

markdown

```
```{r}
dataset <- read.csv(
  "https://raw.githubusercontent.com/japhet125/r-workflow-assignment/main/
  stringsAsFactors = FALSE
)
```

```
head(dataset)
```

```
```
```

⚠️ Notice:

- Triple backticks
- {r} is lowercase
- Straight quotes only
- Code is isolated from text

---

 Copy code

🔧 Step 3: Make sure your headers are real Markdown headers

Your sections should look like this:

```
markdown
```

```
Overview
```

```
(text)
```

```
Data Transformation
```

```
(text)
```

```
Conclusion
```

```
(text)
```

RPubs needs the ## to render sections correctly.

---

✅ What your FINAL Rmd should roughly look like

```
markdown
```

```

```

```
title: "NYC Open Data – R Workflow"
```

```
author: "Guibril Ramde"
```

```
date: "2026-01-25"
```

```
output: html_document
```

```

```

## ## Overview

This project analyzes housing-related data obtained from NYC Open Data.

Source: <https://opendata.cityofnewyork.us/>

## ## Data Transformation

The dataset was cleaned and reduced to a subset of meaningful columns.

```
```{r}
dataset <- read.csv(
  "https://raw.githubusercontent.com/japhet125/r-workflow-assignment/main/",
  stringsAsFactors = FALSE
)

head(dataset)
```

Conclusion

Future work could extend this analysis with additional variables or time periods.

yaml

✅ How to confirm it's fixed

1. **Knit** the document in RStudio
2. Upload to **RPubs again**
3. Open the RPubs link
4. You should see:
 - Output of `head(dataset)`
 - No errors
 - Working data load

🎯 What your professor will see now

- ✓ Proper R Markdown structure
- ✓ Executable code
- ✓ Reproducible data access

✓ Clean narrative sections

This will **fully satisfy the reproducibility requirement**.

If you want, paste your **updated Rmd** here and I'll do a final check bef



```
title: "NYC Open Data – R Workflow"  
author: "Guibril Ramde"  
date: "2026-01-25"  
output: html_document
```

Overview

This project analyzes housing-related data obtained from NYC Open Data. The dataset includes information on housing programs, population counts, and rental costs, and is used here to demonstrate data loading and transformation in R.

Source: <https://opendata.cityofnewyork.us/>

Data Transformation

The original NYC Open Data file contained a large number of demographic and housing-related variables. To simplify downstream analysis, a subset of relevant columns was selected, including program type, Section 8 status, total families, total population, and average gross rent. Column names were also cleaned to improve readability and consistency.

Conclusion

This analysis prepared a clean and reproducible subset of NYC housing data for future exploration. Future work could examine relationships between housing programs, population size, and rental costs or incorporate additional years of data for trend analysis.

```
{r}  
dataset <- read.csv(  
  "https://raw.githubusercontent.com/japhet/  
  stringsAsFactors = FALSE  
)  
  
head(dataset)
```

You're **very close** — the content is good 

There's just **one critical formatting issue** that's causing problems on RPubs.



The main problem (this is important)

Your **YAML header** is missing the required `---` lines at the top and bottom.

Without those, R Markdown does **not recognize**:

- the title
- the author
- the date
- the output format

This can break knitting and publishing.

✓ How to FIX it (exactly what to change)

At the **VERY TOP** of your `.Rmd` file, add `---` like this:

markdown

```
---
```

```
title: "NYC Open Data – R Workflow"
author: "Guibril Ramde"
date: "2026-01-25"
output: html_document
---
```

Everything else in your file can stay mostly the same.

🔧 One small improvement (recommended, not required)

 Copy code

Move your R code chunk into the “Data Transformation” section, so the narrative and code match.

Here is the clean, corrected version of your file (you can copy–paste this):

markdown

```
---
```

```
title: "NYC Open Data – R Workflow"
author: "Guibril Ramde"
date: "2026-01-25"
output: html_document
---
```

Overview

This project analyzes housing-related data obtained from NYC Open Data. The information on housing programs, population counts, and rental costs, and demonstrate data loading and transformation in R.

Source: <https://opendata.cityofnewyork.us/>

Data Transformation

The original NYC Open Data file contained a large number of demographic and variables. To simplify downstream analysis, a subset of relevant columns were included, including program type, Section 8 status, total families, total population. Column names were also cleaned to improve readability and consistency.

```
```{r}
dataset <- read.csv(
 "https://raw.githubusercontent.com/japhet125/r-workflow-assignment/main/",
 stringsAsFactors = FALSE
)

head(dataset)
```

## Conclusion

This analysis prepared a clean and reproducible subset of NYC housing data for future exploration.

Future work could examine relationships between housing programs, population size, and rental costs

or incorporate additional years of data for trend analysis.

yaml

---

## What to do next (final steps)

1. Save the ` `.Rmd`
2. Click \*\*Knit\*\* in RStudio
3. Make sure it renders without errors
4. Publish again to \*\*RPubs\*\*
5. Submit the new RPubs link

---

## What your professor will see now

- ✓ Proper YAML header
- ✓ Clear section headers
- ✓ Executable R code
- ✓ Reproducible data loading
- ✓ Clean narrative

This \*\*fully satisfies the assignment requirements\*\*.

If you want, paste the \*\*RPubs link\*\* here after re-publishing and I'll do



---

```
title: "NYC Open Data – R Workflow"
author: "Guibril Ramde"
date: "2026-01-25"
output: html_document
```

--

## ## Overview

This project analyzes housing-related data obtained from NYC Open Data. The dataset includes information on housing programs, population counts, and rental costs, and is used here to demonstrate data loading and transformation in R.  
Source: <https://opendata.cityofnewyork.us/>

## ## Data Transformation

The original NYC Open Data file contained a large number of demographic and housing-related variables. To simplify downstream analysis, a subset of relevant columns was selected, including program type, Section 8 status, total families, total population, and average gross rent. Column names were also cleaned to improve readability and consistency.

```
{r}
dataset <- read.csv(
 "https://raw.githubusercontent.com/japhet/
 stringsAsFactors = FALSE
)

head(dataset)
```

## ## Conclusion

This analysis prepared a clean and reproducible subset of NYC housing data for future exploration.

Future work could examine relationships between housing programs, population size, and rental costs or incorporate additional years of data for trend analysis.

You're **one character away** from perfection 😊

There's a tiny but critical typo in your YAML header.

---



### The issue

You closed the YAML block with **two dashes** instead of **three**:

markdown

--

R Markdown requires exactly **three dashes** to close the YAML header.

Copy code

---

### ✓ The fix (this is all you need to change)

Replace this:

markdown

--

with this:

Copy code

markdown

----

 Copy code

## Corrected header (copy-paste this)

markdown

```

```

```
title: "NYC Open Data – R Workflow"
author: "Guibril Ramde"
date: "2026-01-25"
output: html_document
```

```

```

Everything below the header is already correct — your sections, narrative, and code chunk are all fine 

---

## What to do next

 Copy code

1. Fix the -- → ---
  2. Save the file
  3. Click Knit
  4. Re-publish to RPubs
- 

## Final reassurance

Once that dash issue is fixed:

- Your document will knit
- The GitHub data link will work
- Your workflow meets **reproducibility + structure requirements**
- This is absolutely acceptable for a class submission

If you want, paste the **error message (if any)** after knitting, or the **RPubs link**, and I'll do a final sanity check before you submit.

