

Módulo II

Taller de IA y CD

Maciel-Cruz, EJ

2025-06-12

Contenidos

1	Conceptos Básicos	1
2	Formato Markdown.	5
3	Esto es un encabezado tipo 1	5
3.1	Esto es un encabezado tipo 2	5
3.1.1	Esto es un encabezado tipo 3	5
4	Generalidades de Bases de Datos	7
4.1	Nombrar archivos	7
4.2	Extensiones de archivos	8
4.3	Tipos y Estructuras de Datos	9
5	Ambiente de R	15
5.1	Paneles de trabajo	15
5.2	Cinta de utilidades	16
5.3	Creación de proyectos y archivos	16

1 Conceptos Básicos

1. Programación

La programación se puede definir como un conjunto de instrucciones para crear una función o programa. Básicamente, es decirle a la computadora, mediante un lenguaje específico, que realice una tarea. La programación puede ser tan básica como lo siguiente:

```
diez <- 5+5  
print(diez)
```

```
[1] 10
```

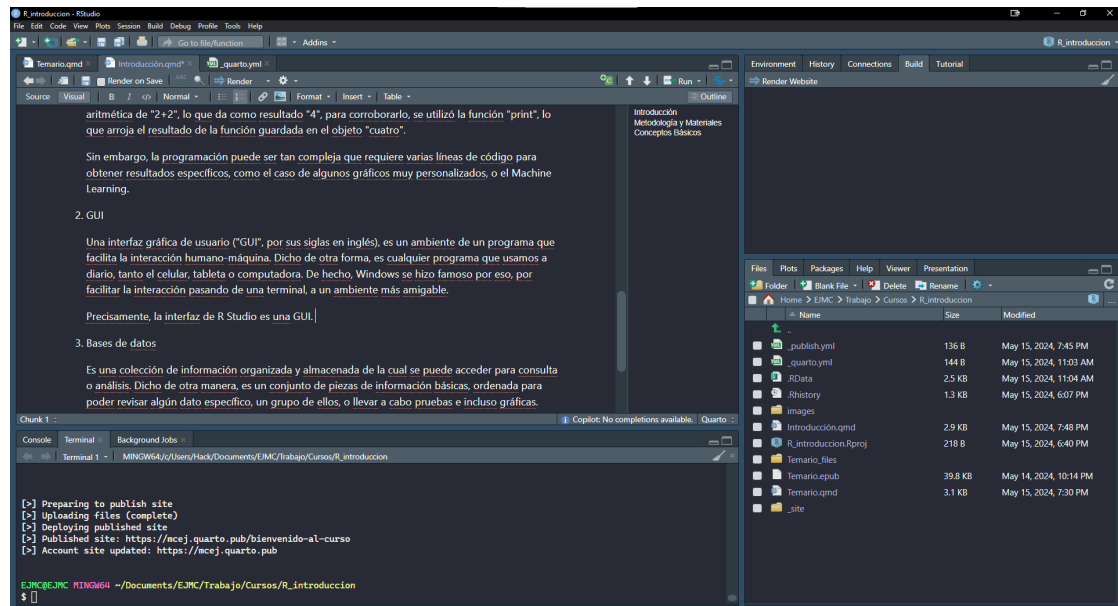
En el código previo, se generó una variable (“diez”), que tiene como función generar la operación aritmética de “5+5”, lo que da como resultado “10”, para corroborarlo, se utilizó la función “print”, lo que arroja el resultado de la función guardada en el objeto “diez”.

Sin embargo, la programación puede ser tan compleja que requiere varias líneas de código para obtener resultados específicos, como el caso de algunos gráficos muy personalizados, o el Machine Learning.

2. GUI

Una interfaz gráfica de usuario (GUI, por sus siglas en inglés) es un entorno visual que permite a las personas interactuar con un programa de manera sencilla y amigable, utilizando elementos gráficos como ventanas, iconos y menús. Este tipo de interfaz facilita el uso de computadoras y dispositivos, ya que no requiere recordar comandos de texto, y es la forma en que la mayoría de los usuarios interactúan con sistemas operativos como Windows, macOS o aplicaciones como R Studio.

Precisamente, la interfaz de R Studio es una GUI.



De hecho, utilizo una GUI (R Studio y VSCode), para generar el curso y poder publicarlo en línea. La parte inferior, que dice “Terminal”, es lo que habitualmente usamos en Linux para introducir comandos.

3. Bases de datos

Una base de datos es un conjunto de información organizada de manera estructurada, que permite almacenar, consultar y analizar datos de forma eficiente. En otras palabras, es una colección de datos relacionados que pueden ser revisados, filtrados o utilizados para realizar análisis y generar reportes o gráficos, dándole un sentido específico para resolver algún problema.

base_cancer.csv - Excel															Enc Jonathan Maciel Cruz														
ArchivosInicioInsertarDiagramaDisposición de páginaFormulasDatosRevisarVistaProgramadorKutoolsKutools PlusAyudaFórmulas ¿Qué desea hacer?																													
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														
Inicio															Inicio														

En la imagen anterior, observamos una base de datos clásica, ordenada y estructurada, las que trabajamos a diario en una GUI llamada Excel.

base_cancer_corregida.csv: Bloc de notas

Archivo Edición Formato Ver Ayuda

folio;estado;edad;peso;altura;subtipo_mol;BRCA1_SNV1;BRCA1_SNV2;BRCA2_SNV1;BRCA2_SNV2;proteina_reactiva;VSG;antigeno_embri;C3;C4;hipotir;AHF_ler_grado;tabaco;alcohol;drogas

448;Aguascalientes;31;71.1193967;1.890186084;LA;VV;MW;VW;7.337;67.004;0.754;142.447;135.982;si;1;no;1;1
565;Aguascalientes;28;91.02203611;1.615018757;LB;MW;MW;VW;3.711;48.661;8.225;148.442;133.453;si;2;si;1;1
1363;Aguascalientes;23;51.09392423;1.62042818;VV;MW;MW;VW;8.747;55.072;2.815;125.69;135.849;no;2;si;1;1
1572;Aguascalientes;34;65.69923727;1.64193918;LA;VV;VW;VW;1.211;52.336;1.949;162.25;162.545;no;0;no;1;1
1984;Aguascalientes;36;80.42910993;1.842593103;LA;MW;MW;MW;5.526;20.345;1.721;89.146;114.931;no;2;si;1;1
2533;Aguascalientes;;93.59057194;1.669600162;MW;VW;MW;VW;9.551;15.801;9.991;129.284;90.294;no;2;si;1;1
2600;Aguascalientes;79;68.67401742;1.857617425;TN;MW;VW;VW;4.368;25.018;8.739;144.763;163.723;no;1;si;1;1
2612;Aguascalientes;;68.70960428;1.541237551;LB;VW;VW;MW;7.301;47.416;7.962;86.29;132.364;no;2;si;1;1
2627;Aguascalientes;52;71.59445329;1.842028032;LA;VV;VW;VW;5.629;36.37;5.949;110.781;155.778;no;1;no;1;1
399;Baja California;75;74.14748217;1.588659104;LA;MW;MW;VW;5.947;48.793;0.822;136.395;86.447;si;2;si;1;1
962;Baja California;78;55.17659637;1.80046031;LA;MW;MW;MW;8.291;31.093;3.024;150.601;82.806;no;2;si;1;1
1337;Baja California;53;54.07501371;1.792658671;LB;VV;MW;5.662;65.01;5.554;147.097;91.429;no;1;no;1;1
1463;Baja California;49;67.88456641;1.714837922;LA;MW;MW;MW;9.21;45.795;7.435;142.005;148.636;no;1;no;1;1
1761;Baja California;58;67.33518477;1.562019933;LA;VW;MW;MW;10.609;64.148;7.9;128.014;126.647;si;2;si;1;1
1883;Baja California;23;72.91519031;1.745186834;TN;VW;VW;MW;8.443;48.625;7.782;130.011;89.872;no;0;si;1;1
2094;Baja California;80;56.25599339;1.7788427;TN;MW;VW;MW;11.662;73.371;9.445;132.664;123.119;no;1;si;1;1
603;Baja California Sur;64;52.2869774;1.868104965;LA;VW;MW;VW;6.264;52.154;7.08;161.595;164.525;no;1;no;1;1
718;Baja California Sur;43;78.1272466;1.754526159;LA;MW;VW;MW;13.133;62.753;4.254;97.023;166.544;no;2;si;1;1
1734;Baja California Sur;70;96.94229406;1.686806006;TN;VW;MW;VW;13.925;19.6;6.604;154.901;152.499;si;2;si;1;1
1847;Baja California Sur;64;64.20286417;1.77788427;TN;MW;MW;VW;3.978;78.584;4.421;128.968;149.734;si;1;no;1;1
2432;Baja California Sur;00;239.2288394;1.75733099;LB;MW;MW;MW;MW;1.006;46.152;8.008;126.132;92.523;no;1;no;1;1
2436;Baja California Sur;44;62.60361165;1.894239218;LA;MW;MW;MW;MW;6.604;26.693;6.284;94.84;142.849;no;2;si;1;1
220;Campeche;;90.54047739;1.641294371;VW;MW;MW;VW;7.203;78.614;9.591;93.076;87.605;si;1;no;1;1
864;Campeche;21;96.52781617;1.745826325;HP;VW;MW;MW;1.989;25.857;8.468;109.27;99.474;no;1;no;1;1
903;Campeche;39;1.566460089;LB;VW;VW;;5.522;50.345;7.529;169.492;162.06;si;2;no;1;1
1676;Campeche;33;1.860142463;TN;VW;MW;MW;12.795;23.23;7.224;115.408;142.325;si;1;si;1;1
2618;Campeche;43;77.36078569;1.525416441;TN;MW;MW;VW;6.2;26.867;4.481;118.663;88.018;no;2;no;1;1
2678;Campeche;49;88.43523716;HP;VW;VW;VW;VW;13.151;64.613;6.47;84.866;152.119;no;0;si;1;1
2931;Campeche;78;71.49028386;1.77514622;LA;MW;VW;MW;VW;9.009;59.235;7.859;111.569;80.282;no;2;si;1;1
750;Chiapas;48;79.8750093;1.595039632;MW;MW;MW;VW;2.819;70.265;0.813;93.646;103.346;no;2;si;1;1
994;Chiapas;27;67.82068164;1.696120975;TN;MW;VW;VW;8.894;78.923;3.267;131.346;138.88;no;0;si;1;1
1344;Chiapas;73;30987641;1.563115971;MW;MW;VW;MW;10.455;12.446;0.567;94.094;153.08;no;1;no;1;1
1649;Chiapas;65;82.92023756;1.640960061;LA;MW;VW;MW;VW;9.479;51.161;7.756;145.446;143.175;no;0;si;1;1
1899;Chiapas;38;50.86305106;1.677747705;LA;MW;MW;MW;MW;5.148;69.943;3.128;157.364;118.876;no;0;no;1;1
2673;Chiapas;25;94.72181539;1.735228632;LB;VW;VW;VW;11.219;23.897;9.073;142.565;149.618;si;2;si;1;1
2803;Chiapas;66;90.49990705;;VW;MW;VW;MW;3.381;60.216;1.078;91.172;108.772;no;1;si;1;1
2829;Chiapas;72;62.15496177;1.506106472;LA;VW;VW;MW;MW;3.544;49.996;0.7;157.408;164.597;no;0;si;1;1
2977;Chiapas;42;76.54731525;1.659513888;LB;VW;MW;VW;MW;10.548;48.271;8.091;124.619;106.694;no;1;no;1;1
57;Chihuahua;73;79.86833973;1.714837922;LA;MW;MW;MW;MW;2.027;63.494;1.74;149.68;132.97;si;1;no;1;1
306;Chihuahua;54;64.33093554;1.81674995;HP;VW;MW;MW;VW;10.456;27.769;2.398;125.876;153.017;no;2;si;1;1
2357;Chihuahua;65;72.47235479;1.868017056;TN;MW;MW;MW;MW;5.357;60.452;7.684;141.545;157.018;si;0;no;1;1
7;Coahuila;67;62.81643369;1.865103045;LB;MW;MW;VW;8.677;46.241;8.622;86.952;86.798;si;0;no;1;1
206;Coahuila;70;52.77309097;1.594687733;HP;MW;MW;MW;12.233;37.189;4.215;167.26;124.691;no;2;si;1;1
237;Coahuila;58;74.37196449;1.757981521;VW;MW;VW;VW;7.915;15.657;8.398;154.991;105.838;no;2;si;1;1
983;Coahuila;76;61.08130703;1.598900536;HP;MW;MW;VW;10.839;14.099;0.832;167.068;83.26;no;1;no;1;1
1008;Coahuila;29;61.29523947;1.50137634;TN;MW;MW;VW;MW;4.584;27.772;6.92;103.88;128.887;no;2;si;1;1
1268;Coahuila;47;63.39642645;1.750929662;MW;MW;VW;MW;8.283;79.622;9.568;167.894;87.551;no;0;no;1;1
1475;Coahuila;67;81.75332718;1.834951478;LB;VW;VW;MW;VW;10.401;8.669;3.072;150.592;169.027;no;2;no;1;1
1493;Coahuila;71;70.56652143;1.545921511;HP;MW;MW;VW;5.898;14.889;3.077;91.304;148.261;no;0;no;1;1
2159;Coahuila;28;51.79239421;1.583564816;HP;VW;VW;MW;10.71;41.205;4.716;164.518;131.687;no;2;si;1;1
2170;Coahuila;66;56.34918798;1.779764647;MW;MW;VW;6.316;62.052;8.713;120.341;100.022;no;2;no;1;1
2335;Coahuila;78;56.35331007;1.842294209;TN;MW;MW;VW;11.661;63.533;6.222;91.229;98.049;si;0;no;1;1

Lo que vemos aquí como un bloque de texto es la misma base de datos, pero en un formato diferente de separación. Este formato permite ahorrar espacio de almacenamiento y facilita su exportación. La base de datos está guardada en un archivo de texto plano y la estamos visualizando con un editor de texto, que también es una GUI.

```
egm09910ELMC -
folio;estado;edad;peso;altura;CM;subtipo.mol;BRCA1_SNV1;BRCA1_SNV2;BRCA2_SNV1;BRCA2_SNV2;proteina_reactiva;V56;antigeno_embri;C3;C4;hipotir;AMF_ler.grado;taba
co;alcohol;drogas
4448;Aguascalientes;31;71;11939967;1.890186084;;LA;VV;VV;VV;7.337;67.004;0.754;142.447;135.982;si;1;no;1;1
965;Aguascalientes;28;91;82203611;1.615010757;;LB;VV;VV;VV;3.711;48.661;0.225;148.442;133.453;si;2;si;1;1
1363;Aguascalientes;23;51;09392423;1.62042818;;VV;VV;VV;VV;8.747;55.072;2.815;125.69;135.849;no;2;si;1;1
1572;Aguascalientes;34;65;69923727;1.64193918;;LA;VV;VV;VV;1.211;52.336;1.949;162.25;162.545;no;0;no;1;1
1984;Aguascalientes;36;80;42910993;1.842593102;;LA;;VV;VV;VV;5.526;28.345;1.721;89.106;114.931;no;2;si;1;1
2533;Aguascalientes;73;59057194;1.669600162;;VV;VV;VV;VV;9.551;15.801;9.991;123.284;90.294;no;2;si;1;1
2668;Aguascalien;79;68.67401742;1.857617425;;TN;VV;VV;VV;VV;4.368;25.018;0.739;144.763;163.723;no;1;si;1;1
2612;Aguascalientes;68.78960428;1.541277551;;LB;VV;VV;VV;VV;7.381;47.416;7.962;86.29;132.364;no;2;si;1;1
2627;Aguascaliente;52;71.59445329;1.842028032;;LA;VV;VV;VV;VV;5.629;36.37;5.949;119.781;165.778;no;1;no;1;1
399;Baja California;75;74.14748217;1.588659104;;LA;VV;VV;VV;VV;5.947;48.793;0.822;136.396;86.447;si;2;si;1;1
962;Baja California;78;55.17659637;1.80046031;;LA;VV;VV;VV;VV;8.291;31.093;3.024;150.601;82.886;no;2;si;1;1
1337;Baja California;53;54.07501371;1.792658671;;LB;;VV;;VV;8.662;65.01;5.584;147.097;91.429;no;1;no;1;1
1463;Baja California;49;67.88456644;1.714837922;;LA;VV;VV;VV;VV;9.21;45.795;7.435;142.805;148.636;no;1;no;1;1
1761;Baja California;58;67.33318477;1.562019933;;LA;VV;VV;VV;VV;10.689;64.140;7.0.128;84;126.647;si;2;si;1;1
1883;Baja California;23;72.91519031;1.745186834;;TN;VV;VV;VV;VV;8.4493;48.625;7.782;138.017;89.872;no;0;si;1;1
1094;Baja California;80;56.25599339;;TN;VV;VV;VV;VV;11.662;73.371;9.445;132.604;133.119;no;1;si;1;1
603;Baja California Sur;64;52.2869774;1.868104965;;LA;VV;VV;VV;VV;6.264;52.154;7.08;161.595;164.525;no;1;no;1;1
1718;Baja California Sur;43;78.1272466;1.754526159;;LA;VV;VV;VV;VV;13.133;62.753;4.254;97.023;166.544;no;2;si;1;1
1734;Baja California Sur;78;96.94229406;1.686806086;;TN;VV;VV;VV;VV;13.925;39.6;6.604;154.901;152.099;si;2;si;1;1
1847;Baja California Sur;64;64.20286417;1.77788427;;TN;VV;VV;VV;VV;3.978;78.584;4.421;128.968;149.734;si;1;no;1;1
2432;Baja California Sur;80;239.2288394;1.75733099;;LB;VV;VV;VV;VV;1.006;46.152;8.088;126.132;92.523;no;1;no;1;1
4436;Baja California Sur;44;62.68361165;1.894239218;;LA;VV;VV;VV;VV;6.604;26.693;6.284;94.84;142.849;no;2;si;1;1
228;Campeche;;90.54047739;1.641294371;;VV;VV;VV;VV;7.203;78.614;9.591;93.076;87.065;si;1;no;1;1
604;Campeche;21;96.52781617;1.745826325;;HP;VV;VV;VV;VV;1.989;25.857;8.468;109.27;99.474;no;1;no;1;1
903;Campeche;39;;1.566460089;;LB;VV;VV;;5.522;50.345;7.529;169.492;162.06;si;2;no;1;1
1676;Campeche;33;1.860142463;;TN;VV;VV;VV;VV;12.795;23.23;7.224;115.408;142.325;si;1;si;1;1
2618;Campeche;43;77.36078569;1.525416441;;TN;VV;VV;;VV;6.2;26.867;4.481;118.663;88.018;no;2;no;1;1
2678;Campeche;49;88.43523716;;HP;VV;VV;VV;VV;13.151;64.613;6.47;84.866;152.119;no;0;si;1;1
2931;Campeche;78;71.49028386;1.775514622;;LA;VV;VV;VV;VV;9.069;59.235;7.859;111.569;80.282;no;2;si;1;1
750;Chiapas;48;79.8758693;1.895039632;;VV;VV;VV;VV;2.819;70.265;0.813;93.648;103.346;no;2;si;1;1
994;Chiapas;27;67.82068164;1.696120975;;TN;VV;VV;VV;VV;8.894;78.923;3.267;121.306;138.88;no;0;si;1;1
1344;Chiapas;;73.30987641;1.563115971;;VV;VV;VV;VV;10.455;12.446;0.567;94.894;153.88;no;1;no;1;1
1649;Chiapas;65;82.92023756;1.640960861;;LA;VV;VV;VV;VV;9.479;51.161;7.756;145.446;143.175;no;0;si;1;1
1899;Chiapas;38;50.86305106;1.677747705;;LA;VV;VV;VV;VV;5.148;69.943;3.128;157.364;118.676;no;0;no;1;1
2673;Chiapas;25;94.72181539;1.735228632;;LB;VV;VV;VV;VV;11.219;23.897;9.073;142.565;149.618;si;2;si;1;1
2883;Chiapas;66;90.49990785;;VV;VV;VV;VV;3.381;60.216;1.078;91.172;108.772;no;1;si;1;1
2829;Chiapas;72;62.15496177;1.506106472;;LA;VV;VV;VV;VV;3.544;49.996;0.7;157.408;164.597;no;0;si;1;1
4977;Chiapas;42;76.54731525;1.659513888;;LB;VV;VV;VV;VV;10.548;48.271;8.091;124.619;106.694;no;1;no;1;1
Encabezado 2
```

2 Formato Markdown.

El formato de archivo Markdown es un lenguaje de marcado ligero que permite crear documentos de texto plano con formato. Es ampliamente utilizado para escribir documentación, notas, blogs y otros tipos de contenido en línea. Markdown utiliza una sintaxis sencilla y fácil de leer, lo que lo hace accesible para usuarios sin experiencia en programación. Lo usaremos durante el taller para crear los documentos y el proyecto final.

Las principales códigos de formato en Markdown incluyen:

- Encabezados: Se crean utilizando el símbolo de número (#) seguido del texto del encabezado. Por ejemplo, # Encabezado 1, ## Encabezado 2, etc.

3 Esto es un encabezado tipo 1

3.1 Esto es un encabezado tipo 2

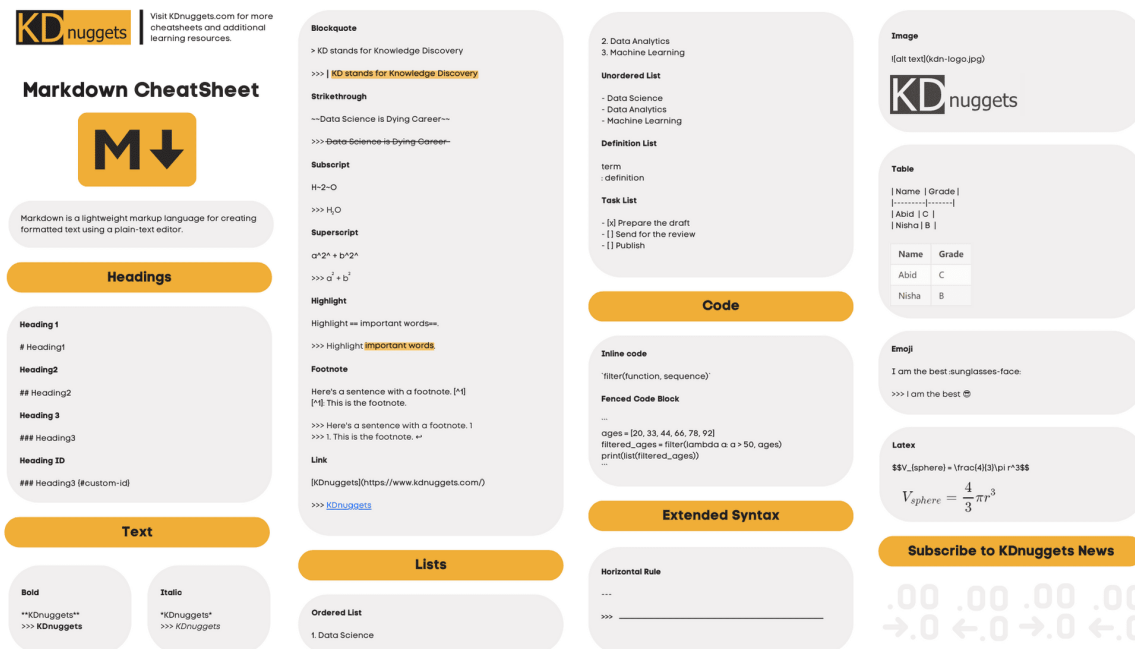
3.1.1 Esto es un encabezado tipo 3

3.1.1.1 Esto es un encabezado tipo 4

- Énfasis: Se utiliza asteriscos (*) o guiones bajos (_) para aplicar énfasis al texto. Por ejemplo, *texto en cursiva* o **texto en negrita**.

- *cursiva*
 - **negritas**
 - ***negritas y cursiva***
- Listas: Se crean utilizando guiones (-) o asteriscos (*) para listas no ordenadas, y números seguidos de un punto (1.) para listas ordenadas. Por ejemplo:
 - Elemento de lista no ordenada
 - * Otro elemento anidado
 - 1. Elemento de lista ordenada
 - 2. Otro elemento
 - 1. Elemento anidado
 - Enlaces: Se crean utilizando corchetes para el texto del enlace y paréntesis para la URL. Por ejemplo, [texto del enlace](#).
 - [Libros gratis](#)

Esos son los más frecuentes, pero dejo una imagen para que vean otros altamente utilizados:



Abid Ali Awan | 2022

Enlace directo a la imagen en la web:

<https://www.kdnuggets.com/wp-content/uploads/markdown-cheatsheet-update.png>

4 Generalidades de Bases de Datos

4.1 Nombrar archivos

Un inconveniente al que nos enfrentamos en R a diferencia de cómo lo hacemos de forma habitual en nuestros archivos de la computadora es el nombramiento de archivos. De forma general, existen las siguientes recomendaciones para R, y su por qué:

1. Usar guión bajo “_”, en vez de espacio “ ”: tanto para los objetos creados en el entorno de R como para los archivos creados que vamos a utilizar en nuestro entorno de trabajo, preferentemente. Esto es debido a que el espacio está reservado como un separador de palabras y objetos, y no es considerado estrictamente como un carácter que utilizamos. Por ejemplo:

```
objeto 1 <- "Objeto 1"
```

Si intentamos correr el texto anterior en nuestro entorno de R, nos arrojaría un error, ya que lo considera como objetos separados, pero si lo hacemos de la siguiente forma:

```
objeto_1 <- "Objeto 1"
```

No se producen errores y se crea el objeto en nuestro ambiente.

Tip

Puedes usar guión bajo, puntos, texto junto, etc., el objetivo es no tener espacios.

1. En los archivos que usemos, como el caso de las bases de datos, se procura usar una nomenclatura similar. Pese a la recomendación, si deseamos utilizar espacios, o tenemos muchos archivos y no los queremos renombrar por cuestiones de tiempo, una forma de solucionar eso es entrecomillar todo el código que tenga dicho espacio, por ejemplo:

```
library("readr") base <- read_csv("~/EJMC/Trabajo/base de datos de prueba.xlsx")
```

El código anterior, todo lo que se encuentra entrecomillado, es la dirección relativa de la base de datos que quiero asignar al objeto llamado base, por lo que le estoy diciendo a mi computadora que, textualmente, busque todo ese segmento de información, porque no está separado.

4.2 Extensiones de archivos

1. CSV

Los archivos .csv son archivos separados por comas (Comma Separated Values), que es un formato especial de documento que permite almacenar grandes cantidades de información en espacios reducidos (pocos megas en comparación con otros formatos como .xlsx), conservando toda su estructura. La desventaja es que es texto plano, es decir, no tiene negritas, cursivas, colores, tamaños de letra ni nada por el estilo. La ventaja, es que se puede manejar cientos de miles de datos fácilmente sin que nuestra computadora termine congelada, dependiendo de la computadora, por supuesto.

El lenguaje R habitualmente maneja este formato de archivos, ya sea separado por comas, tabuladores o algún otro símbolo en especial; sin embargo, también acepta otros formatos, como los de Excel, pero hay que considerar que al momento de importar, se pierde toda la estructura de formato, es decir, colores, tamaños, negritas, cursivas, etc. Por lo que se recomienda habituarse al .csv para fines de ciencia de datos.

2. TSV

El .tsv es similar al formato previo, pero aquí está separado por tabulador, la tecla mencionada previamente. Tiene las mismas ventajas y desventajas que .csv, el por qué usar uno u otro depende de la base de datos en cuestión, si alguna observación cuenta con comas (digamos, 1,234.56), no es recomendable usar .csv., por lo que si se usa .csv, R modificaría la base de datos y no se podría analizar correctamente.

3. XLSX

Este es el formato de excel por excelencia, tiene la característica de que podemos personalizar el tipo de letra, tamaño, colores, etc., pero al cargar a R, este formato se pierde. Para fines de ciencia de datos, se recomienda **NUNCA JAMÁS DE LOS JAMASES EN EL PAÍS DEL NUNCA JAMÁS** (Atentamente: Peter Pan) utilizar colores para dividir información, al menos si piensas hacer análisis con R.

4. TXT

Esto es un archivo de texto plano, es el que se guarda en el bloc de notas, y que se puede abrir con cualquier aplicación similar en cualquier sistema operativo. De hecho, se pueden modificar las extensiones de los .csv y .tsv a .txt y se pueden visualizar de igual forma, siempre y cuando se mantenga la estructura de los separadores.

Tip

El formato .csv es muy flexible, realmente puedes poner separadores como tabular, comas, punto y coma, en R hay una opción para elegir el tipo de separador de dato.

4.3 Tipos y Estructuras de Datos

1. Tipos de datos

En R, se manejan algunos tipos de datos un poco diferentes a lo habitual en estadística, esto con el objeto de hacer operaciones y para ahorrar costos computacionales.

1. Numérico

Es un tipo de dato que contiene número decimales, cualquier número que tenga un punto decimal, debe ser de este tipo.

Hagamos un ejemplo, y creemos los siguientes objetos:

```
n1 <- c(0.1,0.2,1.9,0.5,1.8)
n2 <- c(1,2,3,4,5)
n3 <- c(0,1,1,1,0)
```

Lo anterior, deberá haber creado los objetos “n1”, “n2”, y “n3”, los cuáles contienen los números ingresados.

Además, notarás que existe una “c” previo a la apertura de un paréntesis, esa es la función de combinar o concatenar, es decir, que una serie de datos separados por coma se van a asignar a ese objeto, puede asignar tantos datos como desees.

Tip

En “R” el separador de los datos, ya sea de números o caracteres a combinar, es la coma. Esto también va para separar funciones y argumentos.

Para verificar el contenido, recordemos la función “print”. De igual forma, podemos solo introducir el objeto y oprimir “Ctrl + Enter”, lo que arrojaría el resultado en la consola.

La función “class” nos dice el tipo de clase a la que pertenece algún objeto creado, hagamos el ejemplo:

```
class(n1)
```

```
[1] "numeric"
```

Debería de decir “numeric” en la consola, para cada objeto creado.

Tip

Muchas veces encontrarás un formato llamado “double”, que para fines prácticos del taller es lo mismo que “numeric”.

2. Entero

Es un tipo de dato que contiene números enteros, cualquier número que NO tenga punto decimal, debe de ser de este tipo. Ahora, generemos nuevos objetos, unos de tipo entero, los que desees.

```
e1 <- c(0,0,1,0,1)
e2 <- c(1,2,3,4,5)
e3 <- c(0,1,1,1,0)
```

Ahora, notarás que la clase no se ha modificado desde el primer tipo de dato, es decir, todo nos dice “numeric”, debido a que R almacena cualquier valor numérico en esa clase, tenga o no decimales. El cambio de tipo de dato es útil para algunas funciones, como la notación decimal.

3. Carácter: es cualquier tipo de cadena de texto. Se incluyen símbolos.

```
animales_1 <- c("perros", "gatos", "gallinas")
números_1 <- c(1,2,3)
números_2 <- c("1","2","3")
```

En el ejemplo anterior, creamos tres objetos, “animales_1”, “números_1” y “números_2”. Notarás que unas cadenas de datos contienen comillas, y otras no. Una vez que hayas creado cada uno de los objetos, verifica las clases. Las comillas, para el caso de los números, nos permiten forzar la conversión a carácter. Esto es necesario si los números ingresados corresponde a valores no numéricos, o definiciones previas.

4. Complejo: es cualquier tipo de combinación numérica y una parte imaginaria.

```
complejo <- 5+8i
```

Los tipos de datos complejos se componen de un parte numérica y una imaginaria, que en este caso es el “i”, es poco utilizado en ciencias biológicas. De hecho, no se verá en este curso.

5. Lógico: Es cualquier tipo de operador que entrega dos resultados. Puede ser True o False, 1 y 0, Verdadero o Falso, etc. Se puede asignar y modificar.

```
lógico_1 <- c(T,F,F,T)
logico_1 <- as.logical(c(1,0,0,1))
Lógico_1 <- as.logical(c(0,1,2,3,4))
```

En el objeto “lógico_1”, se genera automáticamente utilizando las iniciales de F o T, pero también se pueden usar las palabras completas en mayúsculas. En el caso de “lógico_2”, nótese que utilizamos la función “as.logical”, lo que hace es forzar la conversión de valores numéricos a lógicos.

Además, nótese que “R” es sensible a los acentos, por lo que ese acento en la “o”, es un objeto diferente. Incluso el objeto “Lógico_1”, contiene valores que van del cero al cuatro, se puede forzar la conversión, pero cualquier valor diferente de “0”, se considera como TRUE.

6. Factores: Realmente no es un tipo de dato de forma específica, sin embargo, presenta una enorme utilidad en el ambiente. En este “tipo de dato”, nos permite categorizar en orden personalizado variables categóricas. Por ejemplo, “bueno”, “regular”, y “malo”. De entrada, se manejaría como carácter, pero podemos modificarlo para que sea factor, es decir, categórico.

```
factores_1 <- as.factor(c(1,2,3))  
factores_2 <- as.factor(c("bueno", "regular", "malo"))
```

Por último, nótese que utilizamos la función “as.factor”, lo que hace forzar la conversión a este tipo de dato. Es importante mencionar que el orden predeterminado es alfanumérico, lo que puede producir problemas de análisis si convertimos a factor una columna sin este método, pero existen varias herramientas para reordenarlos, que se verán más adelante.

Tip

Para fines prácticos del taller, solo nos enfocaremos en “double”, “logical”, “factor”, “character”.

2. Estructuras de datos

Los datos pueden presentar diferentes estructuras, dependiendo si son lineales, o bidimensionales, si contienen un solo tipo de dato (atómicos), o si contienen diferentes tipos.

1. Vectores:

Es la estructura de datos más simple en R, consiste en un conjunto lineal de información de un solo tipo, es decir, solo numérico, o solo carácter, no se debe combinar para que siga siendo un vector.

La función de los vectores es que se pueden aprovechar para realizar la misma operación con el conjunto de los datos.

```
v1 <- c(1:10)  
v2 <- c(2,4,6,8,10)  
v3 <- c(5,10,15,20)  
peso <- c(70,100,80,60)  
altura <- c(1.7,2,1.5,2.1)
```

Los anteriores son vectores numéricos, pero también se pueden hacer con datos de tipo carácter. Con los vectores, se pueden hacer diferentes operaciones matemáticas simultáneas, lo que facilita la generación de datos o nuevas variables.

Hagamos algunas operaciones con los vectores 1 a 3.

```
v1*2
```

```
[1]  2  4  6  8 10 12 14 16 18 20
```

```
v2*2+100
```

```
[1] 104 108 112 116 120
```

```
v3-5
```

```
[1]  0  5 10 15
```

Ahora, obtengamos el imc con la fórmula básica de peso/altura².

```
peso/altura^2
```

```
[1] 24.22145 25.00000 35.55556 13.60544
```

Recordemos el orden de las operaciones en aritmética, estos también se respeta en R. Si imprimimos el resultado, nos debe arrojar el imc para cada dato.

Así mismo, podemos crear un objeto que pase los resultados calculados, esto se logra al crear un objeto, usar el operador de asignación, y crear la aritmética.

```
imc <- peso/altura^2
```

```
imc
```

```
[1] 24.22145 25.00000 35.55556 13.60544
```

Lo anterior nos va a crear un nuevo vector llamado “imc” con el resultado asignado, que es dividir el peso sobre la altura al cuadrado, que es la fórmula del índice de masa corporal. Podemos crear diferentes objetos mediante este formato, es decir, aplicar alguna fórmula, método, conversión, entre otros a algún vector.

Ahora, vamos a crear los siguientes vectores de tipo carácter:

```
v4 <- c("Luis","Juan","María","Lupe")
```

```
v5 <- c("Martínez","Juárez","Domínguez","Jave")
```

El código anterior genera dos vectores diferentes, uno con nombres, y otro con apellidos, en este caso no tendría sentido realizar operaciones aritméticas, incluso no arroja un error. Pero si podemos realizar procesos como juntar ambos vectores para obtener el nombre y apellido, similar a la función de concatenar en Excel.

 Tip

Los vectores deben tener la misma longitud si queremos datos empataados, es decir, observación 1 con observación 2. Si no hubiera congruencia de datos, lo que hace R es reciclar los vectores, o sea, reiniciar el conteo.

Primero vamos a juntar ambos vectores, para eso usamos la función “paste”

```
nombres <- paste(v4,v5)
```

En el código anterior creamos un vector llamado “nombres”, este vector contiene la unión de v4 con v5, en ese orden, para generar un nuevo vector. Por defecto, se utiliza el separador de espacio, si quisiéramos utilizar otro separador, tenemos un argumento llamado “sep”, para que elijamos el símbolo que queramos.

 Tip

En R, un argumento es una opción que modifica la función que estamos utilizando, es decir, podemos hacer unos pequeños ajustes para se comporte como necesitamos. En este caso, “sep” es indicativo de “separate”, es decir, el símbolo que separa la unión de un vector de otro.

2. Matrices

Es una estructura de datos de dos dimensiones (los vectores son tenían una, lo que los hace lineales), por lo que tiene columnas en eje de las y (horizontales), y filas en el eje de las x (verticales).

La clave es que la matriz se puede crear a partir de las filas y columnas indicadas, o se puede transformar acorde a esa razón. Por ejemplo, una matriz de 4 x 3 tendrá cuatro filas y tres columnas, una de 8 x 10, tendrá 8 filas y 10 columnas.

Siempre se cuentan primero las filas, o el eje de las x. Las matrices son especialmente útiles para los datos numéricos, ya que nos permiten hacer transformaciones (aritmética, normalización, estandarización, etc.), reduciendo recursos computacionales.

Vamos a crear la siguiente matriz numérica.

```
m1 <- c(1:50)
dim(m1) <- c(5,10)
print(m1)
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	1	6	11	16	21	26	31	36	41	46
[2,]	2	7	12	17	22	27	32	37	42	47
[3,]	3	8	13	18	23	28	33	38	43	48
[4,]	4	9	14	19	24	29	34	39	44	49
[5,]	5	10	15	20	25	30	35	40	45	50

En el código anterior creamos primeramente un vector llamado “m1”, al que le asignamos un serie de números que van del 1 al 50 (los dos puntos significan que queremos que siga el orden lógico numérico de un inicio a un fin).

Con el objeto creado, vamos a asignarle la dimensión con la función “dim”, y le decimos que sea de 5 x 10, es decir, 5 filas de 10 columnas. Con lo anterior, los números del 1 al 50, se transformarán y se van a reordenar en ese formato, el 1 al 5 en la primera columna, del 6 al 10 en la segunda, y así sucesivamente.

Tip

Si las dimensiones solicitadas no son múltiplo de la cantidad de datos, es decir, si pidiéramos 5 filas con 9 columnas, nos arroja error y nos dice que no coincide la longitud y no se puede crear. Es un método a prueba de error para no perder datos.

La utilidad de las matrices consiste en realizar transformaciones numéricas en nuestras bases de datos, la regla habitual es extraer la matriz de datos en un nuevo objeto, hacer las transformaciones, y llevar a cabo nuestros análisis. La otra opción es introducir la nueva base a la base previa, o reemplazar los valores, como se desee.

Hagamos ejercicios de matrices, creemos dos matrices numéricas, la primera que sea del 1 al 100, en una matrix de 10 x 10; la segunda que sea del 1 al 30, en una matrix de 5 x 6.

3. DataFrame

Son estructuras de datos de dos dimensiones, pero a diferencia de las matrices, los data frame permiten utilizar datos mixtos (combinación de numéricos, lógicos, carácter...), la única condición es que los vectores deben de tener la misma longitud si se crean con R, en caso de que se importen, los datos ausentes se manejan de diferentes formas que se verán más adelante.

De hecho, el data frame, es el tipo de estructura de datos más utilizada en R, y en otros ambientes, nosotros prácticamente solo usamos data frames al trabajar con Excel en el área de ciencias de la salud.

Vamos a crear un data frame (df), de cero.

```

nombre <- c("Luis", "Pedro", "María", "Gloria")
apellido <- c("Gómez", "Fernández", "Cruz", "Pérez")
edad <- c(30,50,23,30)
sobrepeso <- as.logical(c(1,1,0,0))
sexo <- c("Masc","Masc","Fem","Fem")
DM <- c("1","2","0","3")
df1 <- data.frame(nombre, apellido, edad, sobrepeso, sexo, DM)
df1

```

	nombre	apellido	edad	sobrepeso	sexo	DM
1	Luis	Gómez	30	TRUE	Masc	1
2	Pedro	Fernández	50	TRUE	Masc	2
3	María	Cruz	23	FALSE	Fem	0
4	Gloria	Pérez	30	FALSE	Fem	3

Con el código anterior, primero creamos seis vectores que contienen la información de interés, de tipos de datos mixtos, luego, los agregamos todos a un df llamado “df1”, si lo imprimimos, tendremos una base de datos con la información. Además, si observamos en nuestro ambiente, nos refiere el tipo de estructura de dato, que para “df1, es un”data.frame”.



Tip

Para fines prácticos en ciencias biológicas, casi siempre utilizamos dataframes, su abreviatura usual es “df”.

5 Ambiente de R

5.1 Paneles de trabajo

R se divide en cuatro en la interfaz de usuario, esto es útil ya que se puede visualizar todo lo que se requiere y está accesible de forma rápida.

1. **Script:** esquina superior izquierda, aquí escribimos nuestro código.
2. **Entorno:** esquina superior derecha, aquí se encuentra el ambiente de objetos creados, el historial de comandos ingresados, conexiones a sitios web, herramientas de desarrollador (crear páginas web, aplicaciones, libros...), incluso un tutorial para aprender R!
3. **Consola:** esquina inferior izquierda, aquí se encuentra la consola como tal, que es en donde arroja los resultados del código, se imprimen los objetos y nos arroja advertencias y errores si nos equivocamos o nuestro código no está bien escrito (incluso nos dice qué hicimos mal...). Está la terminal, que se basa en el lenguaje de R y funciona similar a la

parte de script, pero con diferente enfoque, aquí se pueden ingresar comandos de administrador y crear otras cosas (como publicar una página web), sin embargo, prácticamente no se usa y no se requiere para empezar a trabajar con R. También están los trabajos en segundo plano o lo que está realizando como procesos.

4. **Salida:** esquina inferior derecha, tiene diferentes partes.
 - **Archivos:** aquí observamos todos los archivos de la carpeta en la que estamos posicionados, esto se puede cambiar según el proyecto, es como un mini explorador de archivos.
 - **Plots:** aquí veremos nuestros gráficos que vamos creando.
 - **Packages,** aquí están las librerías que instalamos y cargamos.
 - **Help,** literalmente, es ayuda, si no sabemos para qué sirve una función, aquí nos arroja un manual de usuario para cada función.
 - **Viewer,** es un explorador parecido a la web, pero con funciones ligeramente diferentes.
 - **Presentation,** nos sirve para visualizar diapositivas de presentación que vayamos creando otro programa, por ejemplo, el presente manual se puede transformar a una presentación estilo power point y se puede visualizar ahí.

5.2 Cinta de utilidades

1. La cinta de utilidad que se encuentra hasta arriba es muy similar a lo que usamos en Office o en otros programas, en cada pestaña encontrarás diferentes utilidades y funciones, como guardar, crear, editar, buscar, personalizar, etc.
2. Debajo de la cinta de utilidades, se encuentran diferentes íconos de acceso rápido, que es crear documento, crear proyecto, abrir archivo, guardar, etc.
3. Debajo de los íconos, se encuentran nuestras pestañas de los archivos abiertos, algo similar a las ventanas del navegador de internet.

5.3 Creación de proyectos y archivos

1. Al momento de trabajar en R, la organización es crucial. De manera general, se recomienda siempre trabajar por proyectos. Los proyectos son carpetas que contienen toda la información necesaria para trabajar en eso, un proyecto, ahí podremos guardar imágenes, bases de datos, libros, archivos de ayuda, además de nuestros scripts.

La ventaja de trabajar por proyectos es que al momento de importar y exportar archivos, es más fácil, y se pueden visualizar en nuestro panel de salida. Otra ventaja de trabajar por proyecto, es que se guarda nuestras librerías cargadas, ambiente, bases cargadas, etc. Lo anterior permite que no mezclamos datos o información de un proyecto a otro, y que no corramos el riesgo de borrar algo que no queríamos.

2. Vamos a crear un proyecto, que tendrá como objetivo recopilar todo lo aprendido en el curso, guardado por secciones o módulos, o cómo deseen (siempre y cuando tenga una jerarquía).
 1. Se sugiere la siguiente estructura básica de almacenamiento:
 - Nombre de la carpeta de trabajo.
 - Nombre del proyecto (p.e. curso_R)
3. Recuerda usar guión bajo, llevar un orden jerárquico de versiones de archivos (ya sea v001, o 30.05.2024), nombres cortos y entendibles, etc.