

In [ ]:

```
# Initialize Otter
import otter
grader = otter.Notebook("hw01_.ipynb")
```

## Tarea 1: Causalidad y Expresiones

Por favor complete este notebook llenando las celdas que se piden. Antes de comenzar, ejecute la celda previa para cargar las pruebas dadas.

### Lecturas recomendadas:

- ¿Qué es la ciencia de datos? (<http://www.inferentialthinking.com/chapters/01/what-is-data-science.html>)
- Causalidad y Experimentos (<http://www.inferentialthinking.com/chapters/02/causality-and-experiments.html>)
- Programando en Python (<http://www.inferentialthinking.com/chapters/03/programming-in-python.html>)

Para todos los problemas en los que deba escribir explicaciones y oraciones, **deberá** proporcionar la respuesta en el espacio designado. Además, a lo largo de esta tarea y de todas las futuras, ¡asegúrese de no reasignar variables en todo el notebook! Por ejemplo, si usa `temperatura` en su respuesta a una pregunta, no la reasigne más adelante. De lo contrario, ¡reprobará las pruebas que creía que estaba aprobando anteriormente!

### Fecha de entrega:

Esta tarea deberá ser entregada \*\* \_\_\_\_ a las \_\_\_\_ \*\*. Si quiere puntos extra, entréguela el (un día antes). Trabajo que se entregue por fuera del plazo no será recibido.

**Nota: Esta tarea tiene pruebas ocultas. Eso significa que aunque los exámenes indiquen que se aprobó al 100%, no significa que su calificación final será del 100%. Realizaremos más pruebas ocultas de corrección una vez que todos entreguen la tarea.**

Evite compartir respuestas o hacer copia con sus compañeros. Establezca en cambio discusiones grupales con otros estudiantes y docentes del curso.

Si desea contar con el tiempo necesario para pedir ayuda o asesoría en caso de estar atascado, se recomienda iniciar temprano con la realización de las tareas. Tenga en cuenta el cronograma del curso y de entregas.

**Q1.** El anuncio de una compañía de alarmas en los Estados Unidos dice,

"Cuando tú te vas de vacaciones, los ladrones aprovechan [...] De acuerdo a estadísticas del FBI, cerca del 25% de los robos ocurren en días festivos."

¿Los datos presentados en el anuncio respaldan la afirmación de que es más probable que los ladrones actúen en días festivos? Por favor explique su respuesta. **(6 Puntos)**

**Nota:** Suponga que cuando se dice "cerca del 25%" es realmente muy cerca de ese valor.

**SOLUTION:** *Escriba su respuesta aquí. Reemplace este texto.*

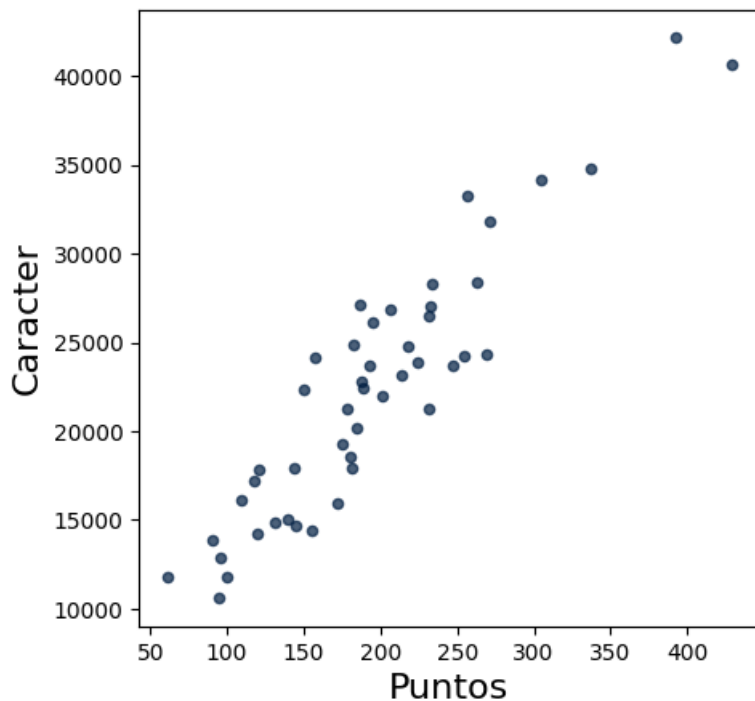
## 2. Caracteres en Little Women

En clase, contamos el número de veces que aparecen los caracteres en cada capítulo del libro clásico [Little Women](https://inferentialthinking.com/chapters/01/3/1/Literary_Characters.html?highlight=little%20women) ([https://inferentialthinking.com/chapters/01/3/1/Literary\\_Characters.html?highlight=little%20women](https://inferentialthinking.com/chapters/01/3/1/Literary_Characters.html?highlight=little%20women)). En ciencias computacionales, la palabra "caracter" también hace referencia a una letra, dígito, espacio o signo de puntuación; cada elemento de un texto. El siguiente código genera una gráfica de dispersión en la cual cada punto corresponde a un capítulo de *Little Women*. La posición horizontal de un punto mide el número de puntos en un capítulo. La posición vertical mide el número total de caracteres.

In [2]:

```
# Solo ejecute esta celda.
# Esta celda contiene código que no ha sido cubierto en este curso,
# sin embargo, usted deberá ser capaz de interpretar la gráfica de dispersión
# que él genera.
from datascience import *
from urllib.request import urlopen
import numpy as np
%matplotlib inline

little_women_url = 'https://www.inferentialthinking.com/data/little_women.txt'
chapters = urlopen(little_women_url).read().decode().split('CHAPTER ')[1:]
text = Table().with_column('Chapters', chapters)
Table().with_columns(
    'Puntos', np.char.count(chapters, '.'),
    'Caracter', text.apply(len, 0)
).scatter(0)
```



**Pregunta 1.** ¿Cerca de cuántos puntos hay en el capítulo con más caracteres? Responda 1,2,3,4, ó 5 al nombre `caracteres_q1` en la siguiente celda. **(4 Puntos)**

1. 250
2. 390
3. 440
4. 32,000
5. 40,000

**Nota:** Si le aparece el error `NameError: name 'grader' is not defined` el calificador de la celda de abajo (y en cualquier otra tarea), ejecute de nuevo la primera celda al inicio de este notebook.

In [3]:

```
# BEGIN SOLUTION NO PROMPT
characters_q1 = 5
# END SOLUTION
""" # BEGIN PROMPT
characters_q1 = ...
""" # END PROMPT
```

Out[3]:

```
' # BEGIN PROMPT\ncharacters_q1 = ...\n'
```

In [ ]:

```
grader.check("q2_1")
```

La celda de arriba verifica que sus respuestas estén en el formato correcto. **Esta prueba no verifica que se haya respondido correctamente**, solo que haya escrito correctamente un número en cada celda de respuesta de selección múltiple.

**Pregunta 2.** ¿Cuál de los siguientes capítulos tiene la mayor cantidad de caracteres por punto? Responda 1, 2, ó 3 en el nombre caracteres\_q2 abajo. **(4 Puntos)**

1. El capítulo con 60 puntos
2. El capítulo con 350 puntos
3. El capítulo con 440 puntos

In [6]:

```
# BEGIN SOLUTION NO PROMPT
characters_q2 = 1
# END SOLUTION
""" # BEGIN PROMPT
characters_q2 = ...
""" # END PROMPT
```

Out [6]:

```
' # BEGIN PROMPT\ncharacters_q2 = ...\n'
```

In [ ]:

```
grader.check("q2_2")
```

De nuevo, la prueba de la celda anterior verifica que sus respuestas estén en el formato correcto, no que haya sido respondido correctamente. Para descubrir más datos interesantes sobre esta gráfica, vaya a la [Section 1.3.2](https://inferentialthinking.com/chapters/01/3/2/Another_Kind_Of_Character.html) ([https://inferentialthinking.com/chapters/01/3/2/Another\\_Kind\\_Of\\_Character.html](https://inferentialthinking.com/chapters/01/3/2/Another_Kind_Of_Character.html)) en el libro guía.

### 3. Nombres y Declaraciones de Asignación

**Pregunta 1.** Cuando ejecuta la siguiente celda, Python produce un mensaje de error.

In [9]:

```
4 = 2 + 2
```

```
Cell In[9], line 1
    4 = 2 + 2
    ^
SyntaxError: cannot assign to literal here. Maybe you meant '==' instead of '='?
```

Escoja la mejor explicación para lo que está mal en el código. Responda 1, 2, 3, ó 4 en nombres\_q1 en la celda de abajo. **(4 Points)**

1. Python es inteligente y sabe que  $4 = 2 + 2$ .
2. En Python, hay una regla en la que el signo `=` debe tener el nombre de una variable a su izquierda y `4` no es el nombre de una variable.
3. Debería ser  $2 + 2 = 4$ .
4. Yo no obtengo ningún mensaje de error. Esta pregunta es capciosa.

In [10]:

```
# BEGIN SOLUTION NO PROMPT
nombres_q1 = 2
# END SOLUTION
""" # BEGIN PROMPT
nombres_q1 = ...
""" # END PROMPT
```

Out[10]:

```
' # BEGIN PROMPT\nnombres_q1 = ...\n'
```

In [ ]:

```
grader.check("q3_1")
```

**Pregunta 2.** Cuando ejecuta la siguiente celda, Python va a arrojar otro mensaje de error.

In [13]:

```
dos = 3
seis = dos mas dos
```

```
Cell In[13], line 2
      seis = dos mas dos
              ^
SyntaxError: invalid syntax
```

Escoja la mejor explicación para lo que está mal con el código y asigne 1, 2, 3 ó 4 en `nombres_q2` abajo para indicar su respuesta- (4 Puntos)

1. La operación `mas` solo aplica a números, no a la palabra "dos".
2. El nombre "dos" no se puede asignar al número 3.
3. Dos más dos es cuatro, no seis.
4. El nombre `mas` no es un operador incorporado; en cambio, la suma usa `+`.

In [14]:

```
# BEGIN SOLUTION NO PROMPT
nombres_q2 = 4
# END SOLUTION
""" # BEGIN PROMPT
nombres_q2 = ...
""" # END PROMPT
```

Out[14]:

```
' # BEGIN PROMPT\nnombres_q2 = ...\n'
```

In [ ]:

```
grader.check("q3_2")
```

**Pregunta 3.** Ejecute la siguiente celda.

In [17]:

```
x = 2
y = 3 * x
x = 4
```

¿Cuál es el valor de `y` luego de correr esta celda? ¿Por qué? Escoja la mejor explicación y asigne 1, 2, 3, ó 4 a `nombres_q3` abajo para indicar su respuesta. (4 Points)

1. `y` es igual a 6, porque el segundo `x = 4` no tiene efecto pues `x` ya había sido definida.
2. `y` es igual a 6, porque `x` era 2 cuando se asignó `y`, y  $3 * 2$  es 6.
3. `y` es igual a 12, porque `x` es 4 y  $3 * 4$  es 12.
4. `y` es igual a 12, porque asignando `x` a 4 va a actualizar `y` a 12 mientras `y` está definida en términos de `x`.

In [18]:

```
# BEGIN SOLUTION NO PROMPT
nombres_q3 = 2
# END SOLUTION
""" # BEGIN PROMPT
nombres_q3 = ...
""" # END PROMPT
```

Out[18]:

```
' # BEGIN PROMPT\nnombres_q3 = ...\n'
```

In [ ]:

```
grader.check("q3_3")
```

Los [datos abiertos](https://www.udea.edu.co/wps/portal/udea/web/inicio/institucional/data-udea/acerca-data-udea/que-data-udea) (<https://www.udea.edu.co/wps/portal/udea/web/inicio/institucional/data-udea/acerca-data-udea/que-data-udea>) de la Universidad de Antioquia provee datos sobre muchos aspectos del campus. Adaptado de su página web, la tabla a continuación muestra el número de egresados en 3 programas entre los años académicos 2009 y 2022

Programa	2009	2022
Biología	76	31
Matemáticas	12	12
Física	24	15
Química	39	22

**Pregunta 1.** Suponga que quiere encontrar la **mayor** diferencia absoluta entre el número de graduados durante esos dos años, entre los tres programas.

En la celda de abajo, calcule este valor y llámelo `mayor_diferencia`. Use una sola expresión (una sola línea de código) para calcular la respuesta. Deje que Python haga toda la aritmética (como sustraer 76 de 31) más que simplificar la expresión usted mismo. La función `abs` toma una entrada numérica y entrega el valor absoluto. La función `max` puede tomar 3 argumentos y devolver el valor máximo entre esos tres números. (5 Puntos)

In [21]:

```
# BEGIN SOLUTION NO PROMPT
mayor_diferencia = max(abs(76 - 31), abs(12 - 12), abs(24 - 15), abs(39 - 22))
# END SOLUTION
""" # BEGIN PROMPT
mayor_diferencia = ...
""" # END PROMPT
```

Out[21]:

```
' # BEGIN PROMPT\nmayor_diferencia = ...\n'
```

In [ ]:

```
grader.check("q4_1")
```

**Pregunta 2.** ¿Cuál de los tres programas tiene la **menor** diferencia absoluta? Asigne `menor_cambio_programa` a 1, 2 ó 3 donde cada número corresponde al siguiente programa:

1. Biología
2. Química
3. Física
4. Matemáticas

Escoja el número que corresponda al programa con el la menor diferencia absoluta. **(4 Puntos)**

*Hint:* Deberá ser capaz de responder aproximadamente con aritmética mental, sin tener que calcular el valor exacto de cada programa.

In [24]:

```
# BEGIN SOLUTION NO PROMPT
menor_cambio_programa = min(abs(76 - 31), abs(12 - 12), abs(24 - 15), abs(39 - 22))
# END SOLUTION
""" # BEGIN PROMPT
menor_cambio_programa = ...
""" # END PROMPT
```

Out[24]:

```
' # BEGIN PROMPT\nmenor_cambio_programa = ...\n'
```

In [ ]:

```
grader.check("q4_2")
```

**Pregunta 3.** Para cada programa, defina el "cambio relativo" como sigue:  $\frac{\text{diferencia absoluta}}{\text{valor en 2009}} * 100$

Llene el código de la siguiente celda de forma tal que `cambio_relativo_fisica`, `cambio_relativo_quimica`, `cambio_relativo_biologia` y `cambio_relativo_matematicas` sean asignados a los cambios relativos de sus respectivos programas. **(5 Puntos)**

Programa	2009	2022
Biología	76	31
Matemáticas	12	12
Física	24	15
Química	39	22

In [27]:

```
# BEGIN SOLUTION NO PROMPT
cambio_relativo_biologia = (abs(31-76) / 76) * 100
cambio_relativo_fisica = (abs(1-12) / 12) * 100
cambio_relativo_matematicas = (abs(24-15) / 24) * 100
cambio_relativo_quimica = (abs(39-22) / 39) * 100

# END SOLUTION
""" # BEGIN PROMPT

cambio_relativo_biologia = (abs(...) / 17) * 100
cambio_relativo_fisica = ...
cambio_relativo_matematicas = ...
cambio_relativo_quimica = ...
cambio_relativo_biologia, cambio_relativo_fisica, cambio_relativo_matematicas, cambio_relativo_quimica
""" # END PROMPT
```

Out[27]:

```
' # BEGIN PROMPT\n\ncambio_relativo_biologia = (abs(...) / 17) * 100\ncambio_relativo_fisica = ...\ncambio_relativo_matematicas = ...\ncambio_relativo_quimica = ...\ncambio_relativo_biologia, cambio_relativo_fisica, cambio_relativo_matematicas, cambio_relativo_quimica\n'
```

In [ ]:

```
grader.check("q4_3")
```

**Pregunta 4.** Asigne `mayor_cambio_rel_programa` a 1, 2, 3 ó 4 donde cada número corresponde con:

1. Biología,
2. Física,
3. Matemáticas,
4. Química

Elija el número que corresponda al programa con el mayor cambio relativo. **(4 Puntos)**

In [34]:

```
# BEGIN SOLUTION NO PROMPT
mayor_cambio_rel_programa=2
# END SOLUTION
""" # BEGIN PROMPT
mayor_cambio_rel_programa = ...
mayor_cambio_rel_programa
""" # END PROMPT
```

Out[34]:

```
' # BEGIN PROMPT\nmayor_cambio_rel_programa = ...\nmayor_cambio_rel_programa\n'
```

In [ ]:

```
grader.check("q4_4")
```

La [miopía \(https://es.wikipedia.org/wiki/Miop%C3%ADa\)](https://es.wikipedia.org/wiki/Miop%C3%ADa) resulta de un número de factores genéticos y ambientales. En 1999, Quinn y colaboradores estudiaron la relación entre la miopía y la iluminación ambiental en la noche (por ejemplo, por las luces nocturnas de las ciudades o la luz de las habitaciones) durante la infancia.

**Pregunta 1.** Los datos reportados en el estudio fueron obtenidos mediante el siguiente procedimiento. "Entre enero y junio de 1998, padres de niños entre los 2 y los 16 años [...] que fueron tratados como pacientes en una clínica pediátrica de oftalmología, llenaron un cuestionario sobre la exposición de luz de los niños antes y después de que hubieran cumplido 2 años. ¿Fue este estudio observacional o fue un experimento controlado? Explique. **(5 Puntos)**

Escriba aquí su respuesta, reemplazando este texto.

**Pregunta 2.** El estudio encontró que de los niños que dormían con la luz de la habitación encendida antes de cumplir los 2 años, el 55% era miope. De los niños que dormían con una lámpara de noche encendida, el 34% eran miopes. De los niños menores de 2 años que dormían en la oscuridad, 10% eran miopes. El estudio concluyó lo siguiente: "La prevalencia de la miopía [...] durante la infancia está fuertemente asociada con la exposición a luz ambiental durante el sueño en la noche, durante los dos primeros años de edad."

**Pregunta 3.** El 13 de mayo de 1999, CNN reportó los resultados de este estudio con el titular: "La luz nocturna puede producir miopía". ¿El estudio original concluye que la luz nocturna puede causar miopía? **(5 Puntos)**

Escriba aquí su respuesta, reemplazando este texto.

**Pregunta 4.** EL párrafo final del reporte de CNN dice que "muchos especialistas" señalaron que el estudio debería considerar la herencia.

La miopía se hereda de los padres a los hijos. Padres miopes tienen mayor probabilidad de tener hijos miopes, y es también más probable que dejen las luces encendidas habitualmente (dado que los padres tienen mala visión). ¿En qué sentido el conocimiento de esta posible relación genética afecta cómo interpretamos los datos del estudio? Explique **(5 Puntos)**

Escriba aquí su respuesta, reemplazando este texto.

## 6. Estudiando a los sobrevivientes

El reverendo Henry Whitehead fue escéptico sobre la conclusión de John Snow sobre el Brote de Broad Street. Luego de que la pandemia de cólera de Broad Street terminó, Whitehead trató de probar que Snow estaba equivocado. (La historia de este evento se detalla [aquí \(https://www.scielo.cl/scielo.php?script=sci\\_arttext&pid=S0716-10182007000400014\)](https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0716-10182007000400014).)

Él notó que Snow había enfocado su análisis casi enteramente en aquellos que fallecieron. Whitehead, por otro lado, investigó los hábitos de bebida de las personas en el área de Broad Street que no habían fallecido en el brote.

¿Cuál es la principal razón por la cual fue importante estudiar este grupo. Asigne 1, 2, ó 3 al nombre `respuesta_supervivientes` abajo. (4 Points)

1. Si Whitehead encontró que muchas personas habían bebido agua del brote de Broad Street y no se contagiaron de cólera, esto puede ser evidencia contra la hipótesis de Snow.
2. Los sobrevivientes pueden proveer información adicional sobre qué más pudo haber causado el cólera, potencialmente apuntando a otra causa.
3. Tras considerar los supervivientes, Whitehead pudo haber identificado una cura para el cólera.

In [38]:

```
# BEGIN SOLUTION NO PROMPT
respuesta_supervivientes = 1

# END SOLUTION
""" # BEGIN PROMPT
respuesta_supervivientes = ...
""" # END PROMPT
```

Out[38]:

```
' # BEGIN PROMPT\nrespuesta_supervivientes = ...\n'
```

**Nota:** Whitehead terminó encontrando pruebas de que el brote de Broad Street jugó un rol clave en la propagación de la enfermedad a personas que vivían cerca. Eventualmente, se convirtió en uno de los mayores defensores de Snow.

In [ ]:

```
grader.check("q6_1")
```

## 7. Políticas y Administratrivia

Esta sección de la tarea es para asegurarnos que leyó la las políticas y sección de preguntas frecuentes del curso.

**Es importante que lea esta sección de la tarea con mucha atención** . Si usted logra terminar esta sección y está seguro de tener todos los recursos correctos configurados, ¡podrá concentrarse en el material real este semestre!

Leyendo la sección de [políticas \(http://data8.org/fa23/policies/\)](http://data8.org/fa23/policies/) y de [preguntas frecuentes \(http://data8.org/fa23/faq/\)](http://data8.org/fa23/faq/) podrá aprobar esta sección con facilidad. Se recomienda leer antes de en efecto ejecutar esta parte de la tarea.

**Pregunta 1.** Suponga que tiene una pregunta respecto a la calificación de algunas de sus tareas que no estaba previamente respondida en la plataforma o en la sección de preguntas frecuentes. ¿A quién contactaría? Asigne `contacto` al número correspondiente con la mejor elección. (4 Points)

1. Los instructores del curso.
2. Publicar una pregunta en la plataforma.
3. Contactar a su profesor del laboratorio.



In [41]:

```
# BEGIN SOLUTION NO PROMPT
contacto = 1

# END SOLUTION
""" # BEGIN PROMPT
contacto = ...
""" # END PROMPT
```

Out[41]:

```
' # BEGIN PROMPT\ncontacto = ...\n'
```

In [ ]:

```
grader.check("q7_1")
```

**Pregunta 2.** Las fechas límites para volver a calificar siempre se publicarán en la plataforma en la que se entregan las notas de las tareas, errores comunes y soluciones. ¿Puede pedir que recalifiquen partes de su tarea luego de que se haya cumplido el plazo de recalificación? Asigne `calificacion` al número correspondiente con la mejor elección. **(4 puntos)**

1. Sí
2. No

In [43]:

```
# BEGIN SOLUTION NO PROMPT
calificacion = 2

# END SOLUTION
""" # BEGIN PROMPT
calificacion = ...
""" # END PROMPT
```

Out[43]:

```
' # BEGIN PROMPT\nncalificacion = ...\n'
```

¡Terminó la tarea 1!

**Información sobre la entrega:** Asegúrese de correr todas las pruebas y verificar que se ejecuten correctamente. Aquí, seleccione **Guardar Notebook** En el menú de **Archivo** y **corra la última celda** para luego darle click en descargar el archivo zip. Con el archivo, vaya a [Gradescope \(https://www.gradescope.com/courses/576157\)](https://www.gradescope.com/courses/576157) y cargue el archivo zip correspondiente a esta tarea. El nombre de esta tarea es "HW 01 Autograder". **Es su responsabilidad asegurarse que su trabajo se guarde antes de correr la última celda** .

Una vez la haya entregado, su tarea en Gradescope se deberá ver como la siguiente imagen si pasó todas las pruebas.

**Nota:** Esta foto es el resultado de someter un atrabajo genérico a Gradescope.



## Estudiantes de Fundamentos de Ciencia de Datos

¡Felicitaciones por terminar la tarea 1!

Close up picture of dog smiling.

Perro de la semana: **Snowball**

## Entrega

Asegúrese que ha ejecutado todas las celdas en su notebook en orden antes de ejecutar la celda de abajo. De esta forma todas las imágenes/gráficas aparecerán en la salida. La celda de abajo va a generar un archivo con extensión .zip para que usted lo entregue. **¡Por favor guarde antes de exportar!**

## Submission

Make sure you have run all cells in your notebook in order before running the cell below, so that all images/graphs appear in the output. The cell below will generate a zip file for you to submit. **Please save before exporting!**

These are some submission instructions.

In [ ]:

```
# Save your notebook first, then run this cell to export your submission.  
grader.export(run_tests=True)
```