1. Load Data

In [1]:
```
import pandas as pd
import geopandas as gpd
import datetime
from ast import literal_eval
```

```
/Users/xuanedx1/opt/anaconda3/lib/python3.7/site-packages/geopandas/
_compat.py:115: UserWarning: The Shapely GEOS version (3.9.1-CAPI-1.
14.2) is incompatible with the GEOS version PyGEOS was compiled with
(3.10.1-CAPI-1.16.0). Conversions between both will be slow.
  shapely_geos_version, geos_capi_version_string
```

In [2]:
```
mobility = pd.read_csv("sg_ca_data.csv", parse_dates=['date_range_star
t'], dtype={0: str, 8: str, 9:str})
```

In [3]:
```
mobility['bucketed_distance_travelled'] = mobility['bucketed_distance_
travelled'].apply(literal_eval)
```

In [4]:
```
mobility.head()
```

Out[4]:

	origin_census_block_group	date_range_start	device_count	distance_traveled_from_home	bu
0	060014075002	2019-12-21 08:00:00+00:00	71	1598.0	
1	060190057022	2019-12-21 08:00:00+00:00	80	3227.0	
2	060210101002	2019-12-21 08:00:00+00:00	101	2405.0	
3	060290060071	2019-12-21 08:00:00+00:00	81	4593.0	
4	060310016011	2019-12-21 08:00:00+00:00	82	4056.0	

In [5]:
```
mobility.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7294981 entries, 0 to 7294980
Data columns (total 10 columns):
 #   Column                           Dtype
---  ------                           -----
 0   origin_census_block_group        object
 1   date_range_start                 datetime64[ns, UTC]
 2   device_count                     int64
 3   distance_traveled_from_home      float64
 4   bucketed_distance_travelled      object
 5   completely_home_device_count     int64
 6   median_percentage_time_home      int64
 7   mean_distance_traveled_from_home float64
 8   county_fips                      object
 9   cbg_fips                         object
dtypes: datetime64[ns, UTC](1), float64(2), int64(3), object(4)
memory usage: 556.6+ MB
```

2. Data Preprocessing

In [6]:
```
# weighted average for bucketed_distance_travelled (by lower bound in distance bin)
distance = [33000, 0, 50000, 5000, 500, 1500, 12000]
mobility['weighted_avg_bucketed_distance_travelled'] = mobility["bucketed_distance_travelled"].apply(lambda x: sum([a*b for a,b in zip(list(x.values()), distance)])/sum(x.values()))
```

In [7]:
```
mobility.head()
```

Out[7]:

	origin_census_block_group	date_range_start	device_count	distance_traveled_from_home	bu
0	060014075002	2019-12-21 08:00:00+00:00	71	1598.0	
1	060190057022	2019-12-21 08:00:00+00:00	80	3227.0	
2	060210101002	2019-12-21 08:00:00+00:00	101	2405.0	
3	060290060071	2019-12-21 08:00:00+00:00	81	4593.0	
4	060310016011	2019-12-21 08:00:00+00:00	82	4056.0	

In [8]:
```
# aggregate data by week and county level
mobility['date_range_start'] = mobility["date_range_start"].apply(lamb
da x: datetime.datetime(year=x.year, month=x.month, day=x.day))
```

In [9]:
```
# aggregate data by week and county level
grouped_mobility = mobility.groupby('county_fips').resample('W', on='d
ate_range_start').sum()
```

In [10]:
```
grouped_mobility
```

Out[10]:

		device_count	distance_traveled_from_home	completely_home_de
county_fips	**date_range_start**			
06001	**2019-12-22**	148322	6296958.0	
	2019-12-29	509476	30129667.0	
	2020-01-05	468948	97359303.0	
	2020-01-12	442203	103785242.0	
	2020-01-19	402867	113142846.0	
...	**...**	...	...	
06115	**2020-10-18**	27823	1139997.0	
	2020-10-25	27213	1153758.0	
	2020-11-01	27177	993137.0	
	2020-11-08	27784	1127616.0	
	2020-11-15	11164	501874.0	

2784 rows × 6 columns

In [11]:
```
# reset index
grouped_mobility = grouped_mobility.reset_index()
grouped_mobility['county_fips'] = grouped_mobility['county_fips'].str[
2:]
```

In [13]:
```
# merge mobility data with geo data
geodata = gpd.read_file("CA_Counties/CA_Counties_TIGER2016.shp")
```

In [14]:
```
geodata.head()
```

Out[14]:

	STATEFP	COUNTYFP	COUNTYNS	GEOID	NAME	NAMELSAD	LSAD	CLASSFP	MTI
0	06	091	00277310	06091	Sierra	Sierra County	06	H1	G4
1	06	067	00277298	06067	Sacramento	Sacramento County	06	H1	G4
2	06	083	00277306	06083	Santa Barbara	Santa Barbara County	06	H1	G4
3	06	009	01675885	06009	Calaveras	Calaveras County	06	H1	G4
4	06	111	00277320	06111	Ventura	Ventura County	06	H1	G4

In [15]:
```
data = grouped_mobility[['county_fips','date_range_start','weighted_avg_bucketed_distance_travelled']].merge(geodata[['COUNTYFP', 'geometry']], how='left', left_on = 'county_fips', right_on='COUNTYFP').sort_values('date_range_start')
```

In [16]:
```
data = data.drop('COUNTYFP', axis=1)
```

In [17]:
```
data
```

Out[17]:

	county_fips	date_range_start	weighted_avg_bucketed_distance_travelled	geometry
0	001	2019-12-22	2.489665e+07	POLYGON ((-13612246.763 4538150.085, -13612346...
1968	083	2019-12-22	7.552061e+06	MULTIPOLYGON (((-13423116.772 4042044.149, -13...
528	023	2019-12-22	2.668767e+06	POLYGON ((-13834943.375 4982761.912, -13834808...
2016	085	2019-12-22	2.562951e+07	POLYGON ((-13585890.862 4467090.861, -13585887...
480	021	2019-12-22	6.906281e+05	POLYGON ((-13680158.035 4814470.618, -13680202...
...	...	...	...	...
1535	063	2020-11-15	1.070617e+06	POLYGON ((-13510515.015 4877236.695, -13510501...
1487	061	2020-11-15	6.883187e+06	POLYGON ((-13476944.153 4722608.150, -13476936...
1439	059	2020-11-15	5.548669e+07	POLYGON ((-13134488.102 3973319.471, -13134662...
2687	111	2020-11-15	1.277050e+07	MULTIPOLYGON (((-13317853.594 3931602.414, -13...
2783	115	2020-11-15	1.755663e+06	POLYGON ((-13536192.850 4739993.980, -13536207...

2784 rows × 4 columns

3. Exploring spatial structure

In [18]:
```
# analysis
import libpysal
from esda.moran import Moran
from esda.moran import Moran_Local
from numpy.random import seed
from libpysal.weights.contiguity import Queen
```

Global and Local Moran I

In [19]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2784 entries, 0 to 2783
Data columns (total 4 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   county_fips                           2784 non-null   objec
t
 1   date_range_start                      2784 non-null   datet
ime64[ns]
 2   weighted_avg_bucketed_distance_travelled  2784 non-null   float
64
 3   geometry                              2784 non-null   geome
try
dtypes: datetime64[ns](1), float64(1), geometry(1), object(1)
memory usage: 108.8+ KB
```

In [20]:
```
idx_df = data.groupby('date_range_start')[['county_fips']].count().cum
sum().reset_index()
```

In [21]:
```
# get global moran I and p-values
import warnings
warnings.filterwarnings("ignore", message="Numerical issues were encou
ntered ")

index = idx_df.county_fips.values
pre_idx = -1
moran_G_raw = [0]*len(index)
moran_L_raw = [0]*len(index)

moran_ = [0]*len(index)
moran_G = [0]*len(index)
p_value_G = [0]*len(index)
moran_L = [0]*len(index)
p_value_L = [0]*len(index)

for i, idx in enumerate(index):
    # Generate W from the GeoDataFrame
    week_df = data.iloc[pre_idx+1:idx,:]
    w = Queen.from_dataframe(week_df, idVariable='county_fips')
    # Row-standardization
    w.transform = 'r'

    # target variable values
    y = week_df['weighted_avg_bucketed_distance_travelled']

    # calculate moran I
    moran = Moran(y, w)

    moran_[i] = moran
    moran_G[i] = moran.I
    p_value_G[i] = moran.p_sim

    moran_L_raw[i] = Moran_Local(y, w)
    moran_L[i] = Moran_Local(y, w).q
    p_value_L[i] = Moran_Local(y, w).p_sim

    # reset left index
    pre_idx = idx
```

```
/Users/xuanedx1/opt/anaconda3/lib/python3.7/site-packages/libpysal/w
eights/_contW_lists.py:31: ShapelyDeprecationWarning: Iteration over
multi-part geometries is deprecated and will be removed in Shapely 2
.0. Use the `geoms` property to access the constituent parts of a mu
lti-part geometry.
  return list(it.chain(*(_get_boundary_points(part.boundary) for par
t in shape)))
/Users/xuanedx1/opt/anaconda3/lib/python3.7/site-packages/libpysal/w
eights/_contW_lists.py:31: ShapelyDeprecationWarning: Iteration over
```

```
eights/_contW_lists.py:31: ShapelyDeprecationWarning: Iteration over
multi-part geometries is deprecated and will be removed in Shapely 2
.0. Use the `geoms` property to access the constituent parts of a mu
lti-part geometry.
  return list(it.chain(*(_get_boundary_points(part.boundary) for par
t in shape)))
/Users/xuanedx1/opt/anaconda3/lib/python3.7/site-packages/libpysal/w
eights/_contW_lists.py:31: ShapelyDeprecationWarning: Iteration over
multi-part geometries is deprecated and will be removed in Shapely 2
.0. Use the `geoms` property to access the constituent parts of a mu
lti-part geometry.
  return list(it.chain(*(_get_boundary_points(part.boundary) for par
t in shape)))
/Users/xuanedx1/opt/anaconda3/lib/python3.7/site-packages/libpysal/w
eights/_contW_lists.py:31: ShapelyDeprecationWarning: Iteration over
multi-part geometries is deprecated and will be removed in Shapely 2
.0. Use the `geoms` property to access the constituent parts of a mu
lti-part geometry.
  return list(it.chain(*(_get_boundary_points(part.boundary) for par
t in shape)))
/Users/xuanedx1/opt/anaconda3/lib/python3.7/site-packages/libpysal/w
eights/_contW_lists.py:31: ShapelyDeprecationWarning: Iteration over
multi-part geometries is deprecated and will be removed in Shapely 2
.0. Use the `geoms` property to access the constituent parts of a mu
lti-part geometry.
  return list(it.chain(*(_get_boundary_points(part.boundary) for par
t in shape)))
```

4. Visualizing and mapping spatial autocorrelation

In [22]:

```
%matplotlib inline
import matplotlib.pyplot as plt
plt.style.use('seaborn-whitegrid')
import seaborn as sns
import numpy as np
```

In [23]:

```
# concat data with week
idx_df['global_moran_I'] = moran_G
idx_df['global_moran_I_p'] = p_value_G
idx_df['local_moran_I'] = moran_L
idx_df['local_moran_I_p'] = p_value_L
```

In [24]:
```
idx_df.head()
```

Out[24]:

	date_range_start	county_fips	global_moran_I	global_moran_I_p	local_moran_I	local_moran
0	2019-12-22	58	0.012896	0.270	[1, 3, 3, 4, 2, 2, 2, 4, 3, 2, 2, 3, 4, 2, 1, ...	[0.103, 0. 0.18, 0. 0.161, 0. 0.
1	2019-12-29	116	0.028421	0.173	[4, 2, 3, 1, 3, 2, 3, 3, 4, 2, 2, 3, 3, 1, 2, ...	[0.155, 0. 0.294, 0. 0.475, 0. (
2	2020-01-05	174	-0.027217	0.486	[2, 3, 3, 2, 3, 2, 4, 3, 3, 3, 4, 3, 2, 4, 2, ...	[0.18, 0. 0.187, 0. 0.071, 0. 0.
3	2020-01-12	232	-0.007273	0.375	[3, 3, 3, 1, 2, 2, 3, 2, 4, 3, 1, 2, 3, 2, 1, ...	[0.397, 0. 0.289, 0. 0.211, 0. (
4	2020-01-19	290	-0.002147	0.345	[3, 3, 4, 4, 4, 2, 3, 3, 3, 1, 2, 1, 2, 4, 3, ...	[0.116, 0. 0.217, 0. 0.467, 0. (

In [25]:
```
idx_df['significance'] = np.where(idx_df['global_moran_I_p'] < 0.05, 'yes', 'no')
idx_df.head()
```

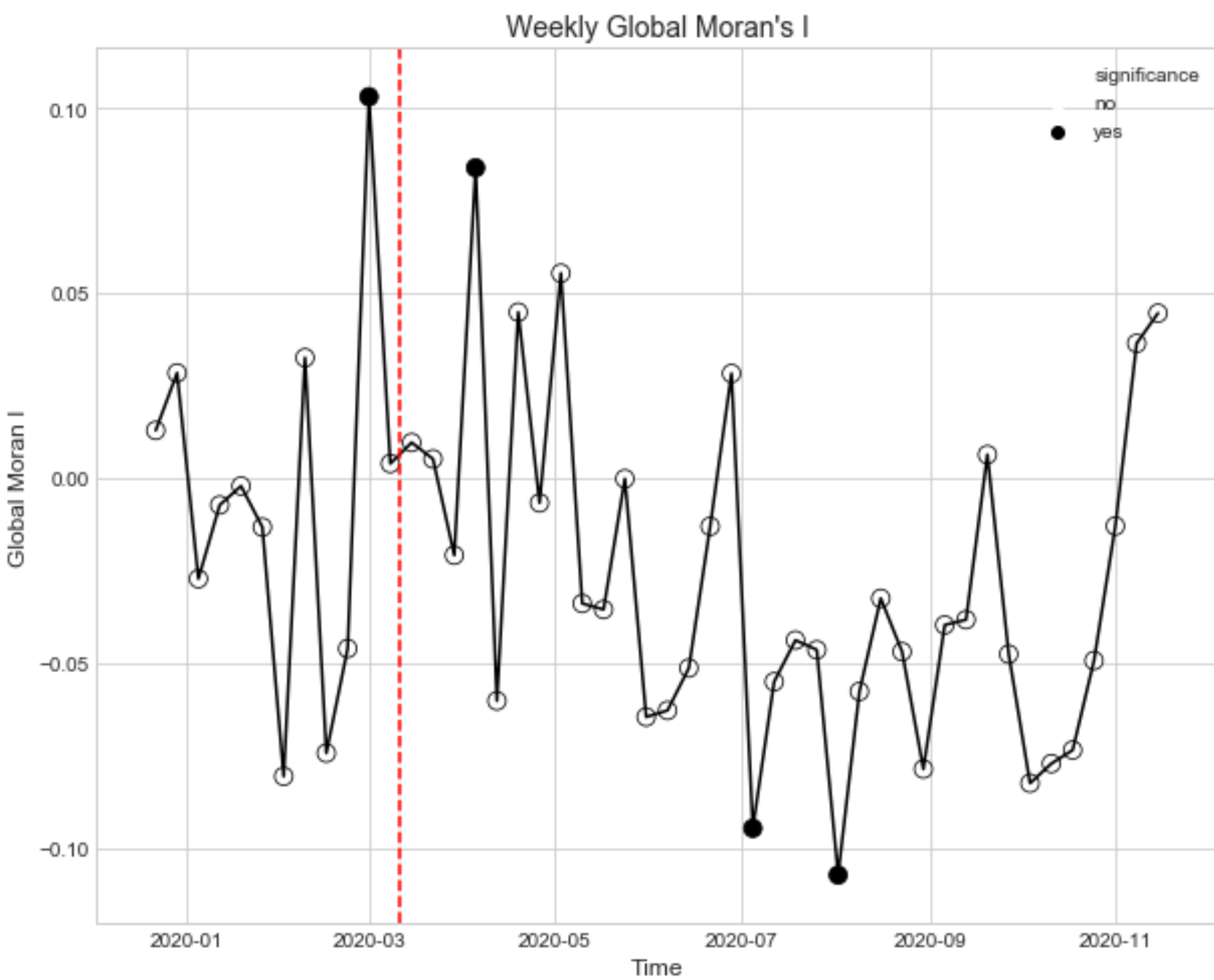
Out[25]:

	date_range_start	county_fips	global_moran_I	global_moran_I_p	local_moran_I	local_moran
0	2019-12-22	58	0.012896	0.270	[1, 3, 3, 4, 2, 2, 2, 4, 3, 2, 2, 3, 4, 2, 1, ...	[0.103, 0. 0.18, 0. 0.161, 0. 0.
1	2019-12-29	116	0.028421	0.173	[4, 2, 3, 1, 3, 2, 3, 3, 4, 2, 2, 3, 3, 1, 2, ...	[0.155, 0. 0.294, 0. 0.475, 0. (
2	2020-01-05	174	-0.027217	0.486	[2, 3, 3, 2, 3, 2, 4, 3, 3, 3, 4, 3, 2, 4, 2, ...	[0.18, 0. 0.187, 0. 0.071, 0. 0.
3	2020-01-12	232	-0.007273	0.375	[3, 3, 3, 1, 2, 2, 3, 2, 4, 3, 1, 2, 3, 2, 1, ...	[0.397, 0. 0.289, 0. 0.211, 0. (
4	2020-01-19	290	-0.002147	0.345	[3, 3, 4, 4, 4, 2, 3, 3, 3, 1, 2, 1, 2, 4, 3, ...	[0.116, 0. 0.217, 0. 0.467, 0. (

Weekly Global Moran's I

In [26]:
```
# global moran I v.s. time
plt.figure(figsize=(10, 8)) # fig size
sns.lineplot(data=idx_df, x='date_range_start', y='global_moran_I', co
lor='black') # lineplot
sns.scatterplot(data=idx_df, x='date_range_start', y='global_moran_I',
hue = 'significance', s=80, palette=dict(no='white',yes='black'), edge
color="black") # add marker
plt.axvline(pd.Timestamp('2020-03-11'), color='r', ls='--') # add vert
ical line
plt.title("Weekly Global Moran's I", fontsize = 14)
plt.xlabel("Time", fontsize = 12)
plt.ylabel("Global Moran I", fontsize = 12)
```

Out[26]:
```
Text(0, 0.5, 'Global Moran I')
```



Observations:

1. There is a decreasing trend in global Moran's I from 03/2020 to 10/2020
2. The 2 greatest and lowest global Moran's I are found to be significant, the rest of the value are insignificant.
3. Overall, the range of Moran's I is small, fluctuating between -0.1 and 0.1
4. The red dash vertical line marked the begining of the pandemic

Interpretations:

1. Overall, the spatial autocorrelations of the average traveled distances of mobile devices is not strong, but comparing the Moran's I across the time we could oberseve the impact of pandemic on the spatial relationship of people's mobility.
2. Before and at the begining of the pandemic, the average traveled distances of mobile devices has a relatively positive spatial autocorrelation at a highest of 0.1, indicating people in counties of similar mibility are closer to each other. However, the spatial autocorrelation went down to negative as the pandemic went on, reaching -0.1 in the summer 2020, showing that people in counties of similar mibility were far away.

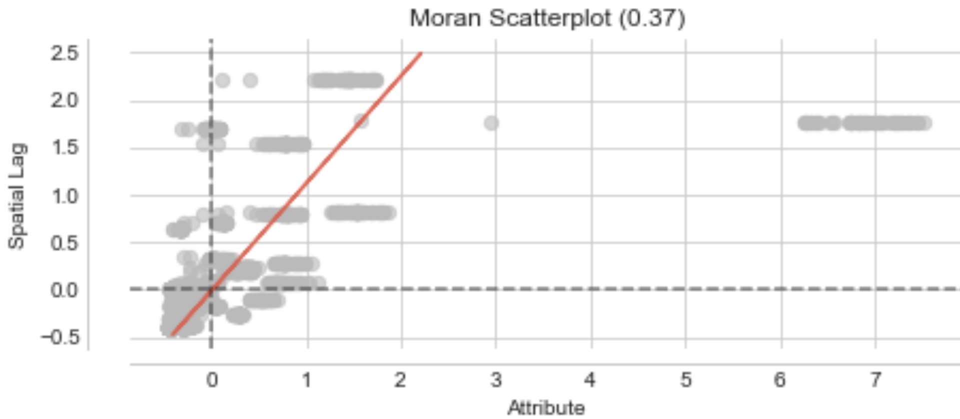
Moran's I Scatterplot for whole dataset

In [27]:
```python
from splot.esda import moran_scatterplot
from splot.esda import plot_local_autocorrelation
```

In [28]:
```python
w = Queen.from_dataframe(data)
moran = Moran(data['weighted_avg_bucketed_distance_travelled'], w)
```

```
/Users/xuanedx1/opt/anaconda3/lib/python3.7/site-packages/libpysal/w
eights/_contW_lists.py:31: ShapelyDeprecationWarning: Iteration over
multi-part geometries is deprecated and will be removed in Shapely 2
.0. Use the `geoms` property to access the constituent parts of a mu
lti-part geometry.
  return list(it.chain(*(_get_boundary_points(part.boundary) for par
t in shape)))
```

In [29]:
```
fig, ax = moran_scatterplot(moran, aspect_equal=True)
plt.figsize=(30,30)
plt.show()
```

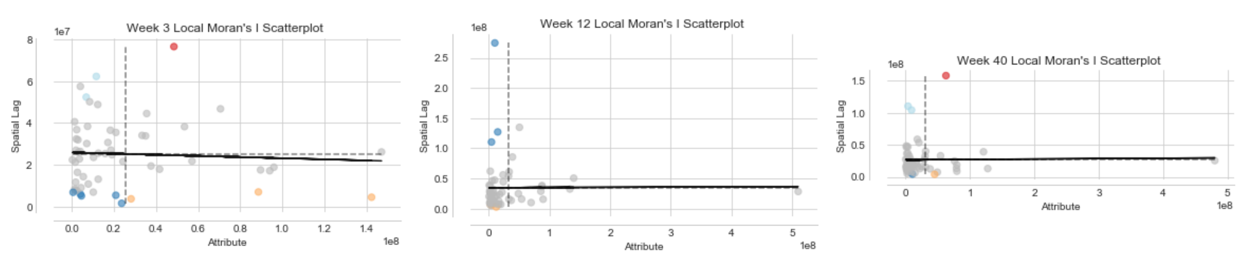


Local Moran's I for Week3, 12, 40

In [30]:
```
# 3 subplots for Local Moran’s (LISA) for Week 3, Week 12, and Week 40
fig, axs = plt.subplots(1, 3, figsize=(20,15),subplot_kw={'aspect': 'equal'})
moran_scatterplot(moran_L_raw[2], p=0.05, zstandard=False, ax=axs[0])
#week 3
moran_scatterplot(moran_L_raw[11], p=0.05, zstandard=False, ax=axs[1])
# week 12
moran_scatterplot(moran_L_raw[39], p=0.05, zstandard=False, ax=axs[2])
# week 40

# set title
axs[0].set_title("Week 3 Local Moran's I Scatterplot ")
axs[1].set_title("Week 12 Local Moran's I Scatterplot ")
axs[2].set_title("Week 40 Local Moran's I Scatterplot ")
```

Out[30]: Text(0.5, 1.0, "Week 40 Local Moran's I Scatterplot ")



Week 3:

1. 1 hotspot county (red) in High-High local spatial autocorrelation zone(upper right), i.e. high mobility county surrouned by high mobility counties
2. 4 coldspot county (blue) in Low-Low local spatial autocorrelation zone(lower left), i.e. low mobility county surrouned by low mobility counties
3. More counties Distributed in Low-High, Low-Low zone, showing many low mobility counties are roughly equally surrounded by high and low mobility counties, i.e. mobility of neighbor counties are dissimilar

Week 12:

1. no hotspot, 2 coldspot.
2. most counties in Low-Low local spatial autocorrelation zone, i.e. mobility of neighbor counties are highly similar in that all are low.

Week 40:

1. 1 hotspot, 1-2 coldspot.
2. most counties in Low-Low local spatial autocorrelation zone, but a hotspot reappear. i.e. most counties has similar low mobility, 1 counties reapprear high spatial autocorrelation in mobility

Overtime, the mobility in California Counties tend to decrease over 2020, however, a few county revive it mobility to high level at the end of the year.

In [31]:
```
!jupyter nbconvert --to pdf --no-input Wenxuan_Zhang_Lab_test.ipynb
```

```
[NbConvertApp] Converting notebook Wenxuan_Zhang_Lab_test.ipynb to p
df
[NbConvertApp] Support files will be in Wenxuan_Zhang_Lab_test_files
/
[NbConvertApp] Making directory ./Wenxuan_Zhang_Lab_test_files
[NbConvertApp] Writing 59521 bytes to ./notebook.tex
[NbConvertApp] Building PDF
[NbConvertApp] Running xelatex 3 times: ['xelatex', './notebook.tex'
, '-quiet']
[NbConvertApp] CRITICAL | xelatex failed: ['xelatex', './notebook.te
x', '-quiet']
This is XeTeX, Version 3.141592653-2.6-0.999993 (TeX Live 2021) (pre
loaded format=xelatex)
 restricted \write18 enabled.
entering extended mode
(./notebook.tex
LaTeX2e <2021-11-15> patch level 1
L3 programming layer <2021-11-22>
(/Users/xuanedx1/Library/TinyTeX/texmf-dist/tex/latex/base/article.c
ls
```

```
convert/exporters/exporter.py", line 197, in from_file
    return self.from_notebook_node(nbformat.read(file_stream, as_ver
sion=4), resources=resources, **kw)
  File "/Users/xuanedx1/opt/anaconda3/lib/python3.7/site-packages/nb
convert/exporters/pdf.py", line 185, in from_notebook_node
    self.run_latex(tex_file)
  File "/Users/xuanedx1/opt/anaconda3/lib/python3.7/site-packages/nb
convert/exporters/pdf.py", line 156, in run_latex
    self.latex_count, log_error, raise_on_failure)
  File "/Users/xuanedx1/opt/anaconda3/lib/python3.7/site-packages/nb
convert/exporters/pdf.py", line 145, in run_command
    command=command, output=out))
nbconvert.exporters.pdf.LatexFailed: PDF creating failed, captured l
atex output:
Failed to run "['xelatex', './notebook.tex', '-quiet']" command:
This is XeTeX, Version 3.141592653-2.6-0.999993 (TeX Live 2021) (pre
loaded format=xelatex)
 restricted \write18 enabled.
entering extended mode
(./notebook.tex
LaTeX2e <2021-11-15> patch level 1
L3 programming layer <2021-11-22>
(/Users/xuanedx1/Library/TinyTeX/texmf-dist/tex/latex/base/article.c
ls
Document Class: article 2021/10/04 v1.4n Standard LaTeX document cla
ss
(/Users/xuanedx1/Library/TinyTeX/texmf-dist/tex/latex/base/size11.cl
o))

! LaTeX Error: File `tcolorbox.sty' not found.

Type X to quit or <RETURN> to proceed,
or enter new name. (Default extension: sty)

Enter file name:
! Emergency stop.
<read *>

l.4     \usepackage
                   {parskip} % Stop auto-indenting (to mimic markdow
n behavi...

No pages of output.
Transcript written on notebook.log.
```

In [ ]: