Mayfly

Reproducible Research in Minutes

python™

Dropbox

OSDC PIRE

- Reproducible Research is the new paradigm
- And more graphics should be interactive
- So, why don't people do it?

- Never learned how
- It's a hassle
- No consistent platform

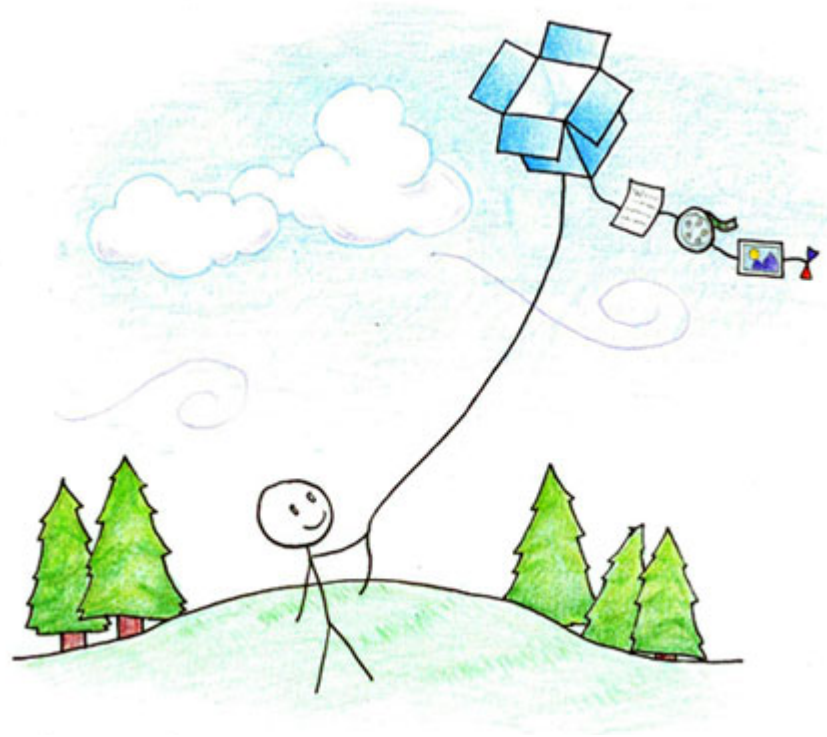
- So:
- Create platform for reproducible research
- Hosted by OSDC

Step 1: Configure Environment

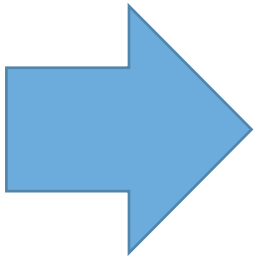
Mayfly

- IPython
- R
- Dropbox
- Dropbox Publisher
- Configures Dropbox

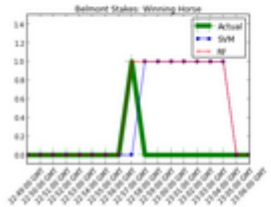
Step 2: Research!



Mayfly

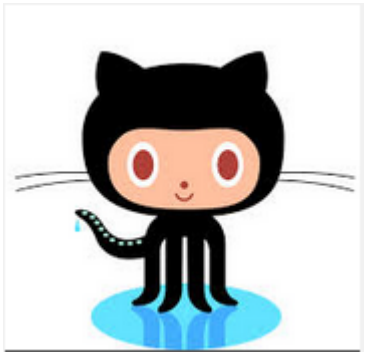


```
for eventname in eventfiles.keys():
    event = eventdata[eventname]
    xdata = event['xdata']
    xlabels = event['xlabels']

    targetevents = event["targets"]

    for targeteventname in targetevents.keys():

        targetevent = targetevents[targeteventname]

        fig, ax = plt.subplots()

        plt.title(eventname + ": " + targeteventname)
        smallerxticks = filter(lambda x: True if x%1==0 else False, xdata)
        pylab.xticks(smallerxticks, [xlabels[x] for x in smallerxticks], rotation=45)

        ax.set_ylim(-0.1, 1.5)
#        ax.set_xlim(0, 175)

        ax.plot(xdata, targetevent["actuals"], color="green", marker='+', lw=7, markersize=20, label="Actual")
        ax.plot(xdata, targetevent["svm"], color="blue", marker='o', markersize=5, label="Svm")
        ax.plot(xdata, targetevent["rf"], color="red", marker='x', markersize=5, label="rf")
        ax.legend()
```



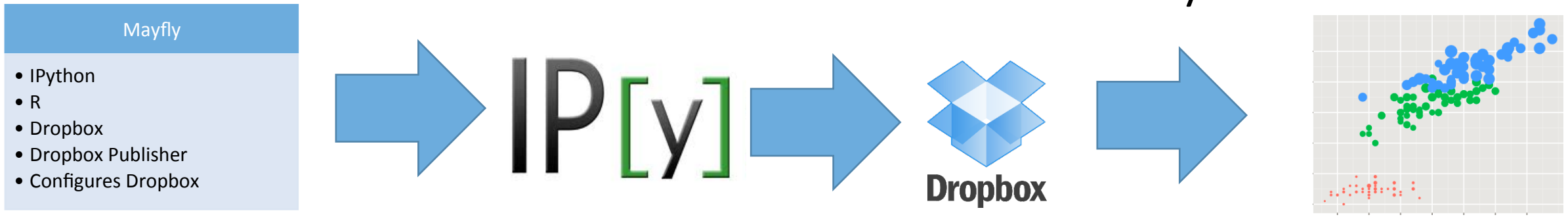
Why is this a good idea?

- Document your research
- Share code, charts, and images
- Host interactive visuals!
- Code is already written



Usage case

- Mary wants to investigate iris flowers
- Needs computational power
- Wants visualizations
- Thinks that reproducible research is good

So,

- Runs setup program
- Points to Python script
- …
- And is given a dropbox link to her report
- Shares with coworkers, gets feedback, spots a coding mistake by her RA…